(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
26 June 2003 (26.06.2003)

PCT

(10) International Publication Number
WO 03/051902 A1

(51) International Patent Classification[7]: C07H 21/04, C12N 15/00, 15/85, 15/86, C07K 5/00, 14/00, A61K 38/00

(21) International Application Number: PCT/US02/40059

(22) International Filing Date:
12 December 2002 (12.12.2002)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

60/340,798	14 December 2001 (14.12.2001)	US
60/365,645	18 March 2002 (18.03.2002)	US
60/367,662	25 March 2002 (25.03.2002)	US
60/379,887	10 May 2002 (10.05.2002)	US
60/384,639	31 May 2002 (31.05.2002)	US

(71) Applicant *(for all designated States except US)*: **INCYTE GENOMICS, INC.** [US/US]; 3160 Porter Drive, Palo Alto, CA 94304 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **BAUGHN, Mariah, R.** [US/US]; 14244 Santiago Road, San Leandro, CA 94577 (US). **BHATIA, Umesh** [US/US]; 5212 Union Avenue, San Jose, CA 95124 (US). **BLAKE, Julie, J.** [US/US]; 3818 Pacheco Street, San Francisco, CA 94116 (US). **BURRILL, John, D.** [US/US]; 2218 Brewster Avenue, Redwood City, CA 94062 (US). **ELLIOTT, Vicki, S.** [US/US]; 3770 Polton Place Way, San Jose, CA 95121 (US). **EMERLING, Brooke, M.** [US/US]; 1735 Woodland Avenue # 71, Palo Alto, CA 94303 (US). **FORSYTHE, Ian, J.** [US/US]; 308 Roble Avenue, Redwood City, CA 94061 (US). **GIETZEN, Kimberly, J.** [US/US]; 691 Los Huecos Drive, San Jose, CA 95123 (US). **GORVAD, Ann, E.** [US/US]; 369 Marie Common, Livermore, CA 94550 (US). **GRIFFIN, Jennifer, A.** [US/US]; 33691 Mello Way, Fremont, CA 94555 (US). **HAFALIA, April, J.A.** [US/US]; 2227 Calle de Primavera, Santa Clara, CA 95054 (US). **HO, Anne** [FR/US]; 1279 Poplar Avenue, #114, Sunnyvale, CA 94086 (US). **JACKSON, Alan, A.** [US/US]; 1541 Elwood Drive, Los Gatos, CA 95032 (US). **JIANG, Xin** [US/US]; 14371 Elva Avenue, Saratoga, CA 95070 (US). **KABLE, Amy, E.** [US/US]; 2345 Polk Street #4, San Francisco, CA 94109 (US). **KEARNEY, Liam** [US/US]; 50 Woodside Avenue, San Francisco, CA 94127 (US). **KHARE, Reena** [US/US]; 12650 Orella Court, Saratoga, CA 95070 (US). **LEE, Ernestine, A.** [US/US]; 20523 Crow Creek Road, Castro Valley, CA 94552 (US). **LEE, Sally** [US/US]; 3643 26th Street, San Francisco, CA 94110 (US). **LU, Dyung Aina, M.** [US/US]; 233 Coy Drive, San Jose, CA 95123 (US). **MARQUIS, Joseph, P.** [US/US]; 4428 Lazy Lane, San Jose, CA 95135 (US). **LEHR-MASON, Patricia, M.** [US/US]; 360 Clarke Lane, Morgan Hill, CA 95014 (US). **RAMKUMAR, Jayalaxmi** [IN/US]; 34359 Maybird Circle, Fremont, CA 94555 (US). **RICHARDSON, Thomas, W.** [US/US]; 616 Canyon Road #107, Redwood City, CA 94062 (US). **SPRAGUE, William, W.** [US/US]; 611 13th Street # C, Sacramento, CA 95814 (US). **TRAN, Uyen, K.** [US/US]; 2638 Mabury Square, San Jose, CA 95133 (US). **CHAWLA, Narinder, K.** [US/US]; 33 Union Square, #712, Union City, CA 94587 (US). **WARREN, Bridget, A.** [US/US]; 1810 S. El Camino Real #B103, Encinitas, CA 94024 (US). **YUE, Henry** [US/US]; 826 Lois Avenue, Sunnyvale, CA 94087 (US). **ZHENG, Wenjin** [CN/US]; 9 Sutter Creek Lane, Mountain View, CA 94043 (US).

(74) Agents: **HAMLET-COX, Diana** et al.; Incyte Genomics, Inc., 3160 Porter Drive, Palo Alto, CA 94304 (US).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, SI, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *with international search report*
— *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments*

[Continued on next page]

(54) Title: NEUROTRANSMISSION-ASSOCIATED PROTEINS

(57) Abstract: Various embodiments of the invention provide human neurotransmission-associated proteins (NTRAN) and polynucleotides which identify and encode NTRAN. Embodiments of the invention also provide expression vectors, host cells, antibodies, agonists, and antagonists. Other embodiments provide methods for diagnosing, treating, or preventing disorders associated with aberrant expression of NTRAN.

WO 03/051902 A1

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

NEUROTRANSMISSION-ASSOCIATED PROTEINS

TECHNICAL FIELD

The invention relates to novel nucleic acids, neurotransmission-associated proteins encoded by these nucleic acids, and to the use of these nucleic acids and proteins in the diagnosis, treatment, and prevention of autoimmune/inflammatory, cardiovascular, neurological, developmental, cell proliferative, transport, psychiatric, metabolic, and endocrine disorders. The invention also relates to the assessment of the effects of exogenous compounds on the expression of nucleic acids and neurotransmission-associated proteins.

BACKGROUND OF THE INVENTION

The human nervous system, which regulates all bodily functions, is composed of the central nervous system (CNS), consisting of the brain and spinal cord, and the peripheral nervous system (PNS), consisting of afferent neural pathways for conducting nerve impulses from sensory organs to the CNS, and efferent neural pathways for conducting motor impulses from the CNS to effector organs. The PNS can be further divided into the somatic nervous system, which regulates voluntary motor activity such as for skeletal muscle, and the autonomic nervous system, which regulates involuntary motor activity for internal organs such as the heart, lungs, and viscera. CNS-associated proteins function in neuronal signaling, cell adhesion, nerve regeneration, axon guidance, neurogenesis, and other processes.

The cerebral cortex or higher brain is the largest structure, consisting of a right and a left hemisphere interconnected by the corpus callosum. The cerebral cortex is involved in sensory, motor, and integrative functions related to perception, voluntary musculoskeletal movements, and the broad range of activities associated with consciousness, language, emotions, and memory. The cerebrum functions in association with the lower centers of the nervous system. The lower areas of the brain such as the medulla, pons, mesencephalon, cerebellum, basal ganglia, substantia nigra, hypothalamus, and thalamus control unconscious activities including arterial pressure and respiration, equilibrium, and feeding reflexes, such as salivation.

The central nervous system (CNS) is composed of more than 100 billion neurons at the spinal cord level, the lower brain level, and the higher brain or cortical level. Neurons transmit electric or chemical signals between cells. The spinal cord, a thin, tubular extension of the central nervous system within the bony spinal canal, contains ascending sensory and descending motor pathways, and is covered by membranes continuous with those of the brainstem and cerebral hemispheres. The

spinal cord contains almost the entire motor output and sensory input systems of the trunk and limbs, and neuronal circuits in the cord also control rhythmic movements, such as walking, and a variety of reflexes. The lower areas of the brain such as the medulla, pons, mesencephalon, cerebellum, basal ganglia, substantia nigra, hypothalamus, and thalamus control unconscious activities including arterial pressure and respiration, equilibrium, and feeding reflexes, such as salivation. Emotions, such as anger, excitement, sexual response, and reaction to pain or pleasure, originate in the lower brain. The cerebral cortex or higher brain is the largest structure, consisting of a right and a left hemisphere interconnected by the corpus callosum. The cerebral cortex is involved in sensory, motor, and integrative functions related to perception, voluntary musculoskeletal movements, and the broad range of activities associated with consciousness, language, emotions, and memory. The cerebrum functions in association with the lower centers of the nervous system.

Nervous system organization and development

A nerve cell (neuron) contains four regions, the cell body, axon, dendrites, and axon terminal. The cell body contains the nucleus and other organelles. The dendrites are processes which extend outward from the cell body and receive signals from sense organs or from the axons of other neurons. These signals are converted to electrical impulses and transmitted to the cell body. The axon, whose size can range from one millimeter to more than one meter, is a single process that conducts the nerve impulse away from the cell body. Cytoskeletal fibers, including microtubules and neurofilaments, run the length of the axon and function in transporting proteins, membrane vesicles, and other macromolecules from the cell body along the axon to the axon terminal. Some axons are surrounded by a myelin sheath made up of membranes from either an oligodendrocyte cell (CNS) or a Schwann cell (PNS). Myelinated axons conduct electrical impulses faster than unmyelinated ones of the same diameter. The axon terminal is at the tip of the axon away from the cell body. (See Lodish, H. et al. (1986) Molecular Cell Biology Scientific American Books New York NY, pp. 715-719.)

CNS-associated proteins have roles in neuronal signaling, cell adhesion, nerve regeneration, axon guidance, neurogenesis, and other functions. Certain CNS-associated proteins form an integral part of a membrane or are attached to a membrane. For example, neural membrane protein 35 (NMP35) is closely associated with neuronal membranes and is known to be highly expressed in the rat adult nervous system (Schweitzer, B. et al. (1998) Mol. Cell. Neurosci. 11:260-273). Synaptophysin (SY) is a major integral membrane protein of small synaptic vesicles. The chromosomal location of SY in human and mouse is on the X chromosome in subbands Xp11.22-p11.23. This region has been implicated in several inherited diseases including Wiskott-Aldrich syndrome, three forms of X-linked hypercalciuric nephrolithiaisis, and the eye disorders retinitis

pigmentosa 2, congenital stationary night blindness, and Aland Island eye disease (Fisher, S.E. et al. (1997) Genomics 45:340-347). Peripherin, or retinal degeneration slow protein (rds), is an integral membrane glycoprotein that is present in the rims of photoreceptor outer segment disks. In mammals, rds is thought to stabilize the disk rim through heterophilic interactions with related nonglycosylated proteins. Rds is a mouse neurological mutation that is characterized by abnormal development of rod and cone photoreceptors followed by their slow degeneration (Kedzierski, W.J. et al. (1999) Neurochem. 72:430-438).

Each of over a trillion neurons in adult humans connects with over a thousand target cells (Tessier-Lavigne, M. et al. (1996) Science 274:1123-1133). These neuronal connections form during embryonic development. Each differentiating neuron sends out an axon tipped at the leading edge by a growth cone. Aided by molecular guidance cues, the growth cone migrates through the embryonic environment to its synaptic target. Progressive axon outgrowth occurs during neural development but not in the mature mammalian CNS. Following CNS injury, expression of growth-inhibiting molecules is enhanced while availability of their growth-promoting counterparts diminishes. Proteins governing developmental axon guidance contribute to the failure of injured central neurons to regenerate. These proteins include Semaphorin3A and the Semaphorin3A receptor proteins neuropilin-1 and plexin-A1 (Pasterkamp, R.J. and J. Verhaagen (2001) Brain Res. Brain Res. Rev. 35:36-54).

Semaphorins function during embryogenesis by providing local signals to specify territories inaccessible to growing axons (Puschel, A.W. et al. (1995) Neuron 14:941-948). They consist of at least 30 different members and are found in vertebrates, invertebrates, and even certain viruses. All semaphorins contain the sema domain which is approximately 500 amino acids in length. Neuropilin, a semaphorin receptor, has been shown to promote neurite outgrowth *in vitro*. The extracellular region of neuropilins consists of three different domains: CUB, discoidin, and MAM domains. The CUB and the MAM motifs of neuropilin have been suggested to have roles in protein-protein interactions and are thought to be involved in the binding of semaphorins through the sema and the C-terminal domains (reviewed in Raper, J.A. (2000) Curr. Opin. Neurobiol. 10:88-94).

The guidance of axons during development involves both positive and negative effects (i.e., chemoattraction and chemorepulsion). The Slit family of proteins have been implicated in promoting axon branching, elongation, and repulsion. Members of the Slit family have been identified in a variety of organisms, including insects, amphibians, birds, rodents and humans (Guthrie, S. (1999) Current Biology 9:R432-R435). Slit proteins are ligands for the repulsive guidance receptor, Roundabout (Robo); however, Slit proteins also cause elongation in some assays. A post-translationally processed form of Slit appears to be the active form of the protein (Guthrie, S. *supra*; and Brose, K. et al. (1999)

Cell 96:795-806).

Axon growth is also guided in part by contact-mediated mechanisms involving cell surface and extracellular matrix (ECM) molecules. Many ECM molecules, including fibronectin, vitronectin, members of the laminin, tenascin, collagen, and thrombospondin families, and a variety of proteoglycans, can act either as promoters or inhibitors of neurite outgrowth and extension (Tessier-Lavigne et al., *supra*). Receptors for ECM molecules include integrins, immunoglobulin superfamily members, and proteoglycans. ECM molecules and their receptors have also been implicated in the adhesion, maintenance, and differentiation of neurons (Reichardt, L.F. et al. (1991) Ann. Rev. Neurosci. 14:531-571). The proteoglycan testican is localized to the post-synaptic area of pyramidal cells of the hippocampus and may play roles in receptor activity, neuromodulation, synaptic plasticity, and neurotransmission (Bonnet, F. et al. (1996) J. Biol. Chem. 271:4373-4380).

Other proteins involved in morphogenetic processes in neural tissues include members of the striatin family such as zinedin and striatin, and members of the sialoadhesins such as myelin-associated glycoprotein (MAG), Schwann cell myelin protein (SMP) and sialoadhesin. Zinedin and striatin are calmodulin-binding, WD repeat proteins primarily expressed in the brain (Castets, F. et al. (2000) J. Biol. Chem. 275:19970-19977). These two proteins share four homologous stretches of amino acids that have been determined to be involved in protein-protein interactions such as caveolin-binding and calmodulin-binding. Both zinedin and striatin have been shown to bind calmodulin in a Ca^{2+}-dependent manner. That striatin is possibly involved in the formation of dendrites has been demonstrated in cultured embryonic motoneurons in which striatin synthesis was specifically blocked. These cells had a reduced number of poorly branched dendrites, whereas the growth of axons appeared normal (Bartoli, M. et al. (1999) J. Neurobiol. 40:234-243). The role of striatin in the control of locomotion has been demonstrated in rats (Bartoli, M. et al. supra). The sialoadhesins are a major subgroup of the I-type lectins which are themselves a subfamily of the lectins. I-type lectins consist mainly of transmembranous glycoproteins belonging to the immunoglobulin superfamily. The sialoadhesins, myelin-associated glycoprotein (MAG), Schwann cell myelin protein (SMP) and sialoadhesin, are expressed in the nervous system and are generally characterized as sialic acid-binding adhesion molecules. MAG is expressed by oligodendrocytes in the CNS in L (large) and S (small) forms that are developmentally regulated such that the L form is expressed predominantly during early stages of myelination and the S form is primarily expressed in adulthood (Tropak, M.B. et al. (1988) Mol. Brain Res. 4:143-155; Pedraza, L. et al. (1991) J. Neurosci. Res. 29:141-148). MAG is also expressed in the Schwann cells in the peripheral nervous system (PNS) and has been implicated in cell recognition mechanisms underlying either the promotion or inhibition of cell adhesion and neurite outgrowth

(Martini, R. (1994) J. Neurocytol. 21:1-28; Bartsch, U. (1996) J. Neurocytol. 25:303-313). SMP is expressed by oligodendrocytes and in Schwann cells, has similar sialic acid-binding properties to MAG and shares 43.5% sequence identity with MAG (Dulac, C. et al. (1992) Neuron 8:323-334; Collins, B. E. et al. (1997) J. Biol. Chem. 272:16889-16895). Sialoadhesin is a membrane glycoprotein originally described as a cell surface receptor expressed by murine macrophages (Crocker, P.R. and Gordon, S. (1986) J. Exp. Med. 164:1862-1875; Crocker, P.R. and Gordon, S. (1989) J. Exp. Med. 169:1333-1346). Expression of sialoadhesin has since been demonstrated in rat macrophages which have been induced by glucocorticoids and can be further enhanced by the addition of interferon β or γ, interleukin 4 or lipopolysaccharides (Van den Berg T.K. et al. (1996) J. Immunol. 157:3130-3138). Expression patterns and binding specificities of sialoadhesin suggest a possible role in regulation of myeloid cell development (Crocker, P.R. et al. (1990) Blood 76:1131-1138; Crocker P.R. et al. (1995) J. Clin. Invest. 95:635-643).

Neurotrophins regulate development, maintenance, and function of vertebrate nervous systems. Neurotrophins activate two different classes of receptors, the Trk family of receptor tyrosine kinases and p75NTR, a member of the TNF receptor superfamily. Through these receptors, neurotrophins activate many signaling pathways, including those mediated by ras and members of the cdc-42/ras/rho G protein families, and by the MAP kinase, PI-3 kinase, and Jun kinase cascades. During development, limiting amounts of neurotrophins function as survival factors to ensure a match between the number of surviving neurons and the requirement for appropriate target innervation. They also regulate cell fate decisions, axon growth, dendrite pruning, the patterning of innervation, and the expression of proteins crucial for normal neuronal function, such as neurotransmitters and ion channels. These proteins also regulate many aspects of neural function. In the mature nervous system, they control synaptic function and synaptic plasticity, while continuing to modulate neuronal survival (Huang, E.J. and L.F. Reichardt (2001) Ann. Rev. Neurosci. 24:677-736). Neuritin is a protein induced by neural activity and by neurotrophins which promote neuritogenesis.

The neurexophilins are neuropeptide-like proteins which are proteolytically processed after synthesis. They are ligands for the neuron-specific cell surface proteins, the α-neurexins. Neurexophilins and neurexins may participate in a neuron signaling pathway (Missler, M. and T.C. Sudhof (1998) J. Neurosci. 18:3630-3638; Missler, M. et al. (1998) J. Biol. Chem. 273:34716-34723). Ninjurin is a neuron cell surface protein which plays a role in cell adhesion and in nerve regeneration following injury. Ninjurin is up-regulated after nerve injury in dorsal root ganglion neurons and in Schwann cells (Araki, T. and J. Milbrandt (1996) Neuron 17:353-361). Ninjurin2 is expressed in mature sensory and enteric neurons and promotes neurite outgrowth. Ninjurin2 is upregulated in

Schwann cells surrounding the distal segment of injured nerve with a time course similar to that of ninjurin1, neural CAM, and L1 (Araki, T. and J. Milbrandt (2000) J. Neurosci. 20:187-195).

Neurexin IV is essential for axonal insulation in the PNS in embryos and larvae. Axonal insulation is of key importance for the proper propagation of action potentials. Caspr, a vertebrate homolog of Neurexin IV -- also named paranodin -- is found in septate-like junctional structures localized to the paranodal region of the nodes of Ranvier, between axons and Schwann cells. Caspr/paranodin is implicated in blood-brain barrier formation, and linkage of neuronal membrane components with the axonal cytoskeletal network (Bellen, H.J. et al. (1998) Trends Neurosci. 21:444-449).

Mammalian Numb is a phosphotyrosine-binding (PTB) domain-containing protein which may be involved in cortical neurogenesis and cell fate decisions in the mammalian nervous system. Numb's binding partner, the LNX protein, contains four PDZ domains and a ring finger domain and may participate in a signaling pathway involving Numb. PDZ domains have been found in proteins which act as adaptors in the assembly of multifunctional protein complexes involved in signaling events at surfaces of cell membranes (Ponting, C.P. (1997) Bioessays 19:469-479). LNX contains a tyrosine phosphorylation site which may be important for the binding of other PTB-containing proteins such as SHC, an adaptor protein which associates with tyrosine-phosphorylated growth factor receptors and downstream effectors (Dho, S.E. et al. (1998) J. Biol. Chem. 273:9179-9187).

Nogo has been identified as a component of the central nervous system myelin that prevents axonal regeneration in adult vertebrates. Cleavage of the Nogo-66 receptor and other glycophosphatidylinositol-linked proteins from axonal surfaces renders neurons insensitive to Nogo-66, facilitating potential recovery from CNS damage (Fournier, A.E. et al. (2001) Nature 409:341-346).

Homeobox transcription factors direct nerve-cell associated tissue patterning and differentiation. The presence and function of these proteins appears to be ubiquitous in nematodes, arthropods, and vertebrates. One example of these proteins is DRG11, a homeobox transcription factor expressed in mammalian sensory neurons, and which appears to be involved in neural crest development (Saito, T. et al. (1995) Mol. Cell Neurosci. 6:280-292). Cutaneous sensory neurons that detect noxious stimuli project to the dorsal horn of the spinal cord, while those innervating muscle stretch receptors project to the ventral horn. DRG11 is required for the formation of spatio-temporally appropriate projections from nociceptive sensory neurons to their central targets in the dorsal horn of the spinal cord (Chen, Z.F. et al. (2001) Neuron 31:59-73).

Synapses

Contact between one neuron and another occurs at a specialized site called the synapse.

Many nervous system functions are regulated by diverse synaptic proteins such as synaptophysin, the synapsins, growth associated protein 43 (GAP-43), SV-2, and p65, which are distributed in subcellular compartments of the synapse. Synaptic terminals also contain many other proteins involved in calcium transport, neurotransmission, signaling, growth, and plasticity. At this site, the axon terminal from one neuron (the presynaptic cell) sends a signal to another neuron (the postsynaptic cell). Synapses may be connected either electrically or chemically. An electrical synapse consists of gap junctions connecting the two neurons, allowing electrical impulses to pass directly from the presynaptic to the postsynaptic cell. In a chemical synapse, the axon terminal of the presynaptic cell contains membrane vesicles containing a particular neurotransmitter molecule. A change in electrical potential at the nerve terminal results in the influx of calcium ions through voltage-gated channels which triggers the release of the neurotransmitter from the synaptic vesicle by exocytosis. The neurotransmitter rapidly diffuses across the synaptic cleft separating the presynaptic nerve cell from the postsynaptic cell. The neurotransmitter then binds receptors and opens transmitter-gated ion channels located in the plasma membrane of the postsynaptic cell, provoking a change in the cell's electrical potential. This change in membrane potential of the postsynaptic cell may serve either to excite or inhibit further transmission of the nerve impulse.

Presynaptic calcium channel activity is modulated by cysteine-string proteins (CSPs). CSPs are secretory vesicle proteins that function in neurotransmission as well as in exocytosis in other cell-types. CSPs belong to the DnaJ/hsp40 (heat shock protein) chaperone family. The effect of CSPs on calcium levels is likely to be downstream of calcium release and is likely to involve exocytosis, possibly in connection with G-proteins (Braun, J.E. et al. (1995) Neuropharmacology 34:1361-9136; Magga, J.M. et al. (2000) Neuron 28:195-204; Dawson-Scully, K. et al. (2000) J. Neurosci. 20:6039-6047; and Chamberlain, L.H. et al. (2001) J. Cell Sci. 114:445-455). Neuregulins (NRGs) mediate between the electrical neural activity and molecular components by regulating the expression of ion channel receptors or transmitter release in synapses. NRGs may also be signaling factors involved in tuning locomotion or other higher functions by coordinating excitatory and inhibitory neurons (Ozaki, M. (2001) Neuroscientist 7:146-154).

N- and P/Q-type Ca^{2+} channels are localized in high density in presynaptic nerve terminals and are crucial elements in neuronal excitation-secretion coupling. In addition to mediating Ca^{2+} entry to initiate transmitter release, they are thought to interact directly with proteins of the synaptic vesicle docking/fusion machinery. N-type and P/Q-type Ca^{2+} channels are colocalized with syntaxin in high-density clusters in nerve terminals. The synaptic protein interaction (synprint) sites in the intracellular loop II-III (LII-III) of both alpha 1B and alpha 1A subunits of N-type and P/Q-type Ca^{2+}

channels bind to syntaxin, SNAP-25, and synaptotagmin. Presynaptic Ca^{2+} channels not only provide the Ca^{2+} signal required by the exocytotic machinery, but also contain structural elements that are integral to vesicle docking, priming, and fusion processes (Catterall, W.A. (1999) Ann. NY Acad. Sci. 868:144-159). Synaptotagmins are a large family of proteins involved in both regulated and constitutive vesicular trafficking. They include a neuronal type (synaptotagmin I-V, X, and XI) and a ubiquitous type (synaptotagmin VI-IX). Ca^{2+}-dependent synaptotagmin activation is involved in neurite outgrowth (Mikoshiba, K. et al. (1999) Chem. Phys. Lipids 98:59-67).

Proteins associated with the membranes of synaptic vesicles include vamp (synaptobrevin), rab3A, synaptophysin, synaptotagmin (p65) and SV2. These membrane proteins function in regulated exocytosis by regulating neurotransmitter uptake, vesicle targeting, and fusion with the presynaptic plasma membrane (Elferink, L.A. and R.H. Scheller (1993) J. Cell Sci. Suppl.17:75-79).

Physophilin, also known as the Ac39 subunit of the V-ATPase, is an oligomeric protein that binds the synaptic vesicle protein synaptophysin, constituting a complex that may form the exocytotic fusion pore. Ac39 is present in a synaptosomal complex which, in addition to synaptophysin, includes the bulk of synaptobrevin II, and subunits c and Ac115 of the V0 sector of the V-ATPase. *In situ* hybridization in rat brain reveals a largely neuronal distribution of Ac39/physophilin mRNA which correlates spatio-temporally with those of subunit c and synaptophysin. Immunohistochemical analysis shows that Ac39/physophilin is mostly concentrated in the neuropil with a pattern identical to subunit A and very similar to synaptophysin. Double-labeling immunofluorescence shows a complete colocalization of Ac39/physophilin with subunit A and a partial colocalization with synaptophysin in the neuropil (Carrion-Vazquez M. et al. (1998) Eur. J. Neurosci.10:1153-1166).

The plasma membrane dopamine transporter (DAT) is essential for the reuptake of released dopamine from the synapse. Uptake of dopamine is temperature- and time-dependent, and is inhibited by a variety of compounds, such as cocaine. DAT-knockout mice have been shown to exhibit extreme hyperactivity and resistance to both cocaine and amphetamine, consistent with the primary action of cocaine on DAT (Giros, B. et al. (1996) Nature 379:606-612). The perturbation of the tightly regulated DAT also predisposes neurons to damage by a variety of insults. Most notable is the selective degeneration of DAT-expressing dopamine nerve terminals in the striatum thought to underlie Parkinson's disease. DAT expression can predict the selective vulnerability of neuronal populations, which suggests that therapeutic strategies aimed at altering DAT function could have significant benefits in a variety of disorders (Gary, W.M. et al. (1999) Trends Pharmacol. Sci. 20:424-429).

43 KD postsynaptic protein or acetylcholine receptor-associated 43 KD protein (RAPSYN) is

thought to play a role in anchoring or stabilizing the nicotinic acetylcholine receptor at synaptic sites. RAPSYN is involved in membrane association and may link the nicotinic acetylcholine receptor to the underlying postsynaptic cytoskeleton (Buckel, A. et al. (1996) Genomics 35:613-616). Neuritin is a protein whose gene is known to be induced by neural activity and by neurotrophins which promote neuritogenesis. Neuraxin is a structural protein of the rat central nervous system that is believed to be immunologically related to microtubule-associated protein 5 (MAP5). Neuraxin is a novel type of neuron-specific protein which is characterized by an unusual amino acid composition, 12 central heptadecarepeats and putative protein and membrane interaction sites. The gene encoding neuraxin is unique in the haploid rat genome and is conserved in higher vertebrates. Neuraxin is implicated in neuronal membrane-microtubule interactions and is expressed throughout the rodent CNS (Rienitz, A. et al. (1989) EMBO J. 8:2879-2888).

Synaptic nuclear envelope-1 (Syne-1) protein is selectively associated with synaptic nuclei and may be involved in the formation or maintenance of nuclear aggregates at the neuromuscular junction. Syne-1 contains multiple spectrin repeats similar to those found in dystrophin and utrophin, as well as a domain homologous to the carboxyl-terminal of Klarsicht, a protein associated with nuclei and required for a subset of nuclear migrations in Drosophila (Apel, E. D. et al. (2000) J. Biol. Chem. 275:31986-31995.).

C1q-Related Factor (CRF) is found to be expressed at highest levels in the brain, particularly in the brainstem. In situ hybridization to mouse brain sections demonstrated that CRF transcripts are most abundant in areas of the nervous system involved in motor function, such as the Purkinje cells of the cerebellum, the accessory olivary nucleus, the pons and the red nucleus. CRF is predicted to encode a polypeptide with a hydrophobic signal sequence, a collagenous region, and a globular domain at the carboxy terminus that shares homology to the C1q signature domain, a subunit of the C1 enzyme complex that activates the serum complement system (Berube, N. G. et al. (1999) Brain Res. Mol. Brain Res. 63:233-240).

Neurotransmitters and neurotransmitter transport proteins

Contact from one neuron to another occurs at a specialized site called the synapse. At this site, the axon terminal from one neuron (the presynaptic cell) sends a signal to another neuron (the postsynaptic cell). Synapses may be connected either electrically or chemically. An electrical synapse consists of gap junctions connecting the two neurons, allowing electrical impulses to pass directly from the presynaptic to the postsynaptic cell. In a chemical synapse, the axon terminal of the presynaptic cell contains membrane vesicles containing a particular neurotransmitter molecule. A change in electrical potential at the nerve terminal results in the influx of calcium ions through voltage-

gated channels which triggers the release of the neurotransmitter from the synaptic vesicle by exocytosis. The neurotransmitter rapidly diffuses across the synaptic cleft separating the presynaptic nerve cell from the postsynaptic cell. The neurotransmitter then binds receptors and opens transmitter-gated ion channels located in the plasma membrane of the postsynaptic cell, provoking a change in the cell's electrical potential. This change in membrane potential of the postsynaptic cell may serve either to excite or inhibit further transmission of the nerve impulse. Presynaptic calcium channel activity is modulated by cysteine-string proteins (CSPs). CSPs are secretory vesicle proteins that function in neurotransmission as well as in exocytosis in other cell-types. CSPs belong to the DnaJ/hsp40 (heat shock protein) chaperone family. The effect of CSPs on calcium levels is likely to be downstream of calcium release and is likely to involve exocytosis, possibly in connection with G-proteins (Braun, J.E. et al. (1995) Neuropharmacology 34:1361-9136; Magga, J.M. et al. (2000) Neuron 28:195-204; Dawson-Scully, K. et al. (2000) J. Neurosci. 20:6039-6047; and Chamberlain, L.H. et al. (2001) J. Cell Sci. 114:445-455).

Neurotransmitters comprise a diverse group of some 30 small molecules which include acetylcholine, monoamines such as serotonin, dopamine, and histamine, and amino acids such as gamma-aminobutyric acid (GABA), glutamate, and aspartate, and neuropeptides such as endorphins and enkephalins (McCance, K.L. and S.E. Huether (1994) PATHOPHYSIOLOGY, The Biologic Basis for Disease in Adults and Children, 2nd edition, Mosby, St. Louis, MO, pp. 403-404). Many of these molecules have more than one function and the effects may be excitatory, e.g. to depolarize the postsynaptic cell plasma membrane and stimulate nerve impulse transmission, or inhibitory, e.g. to hyperpolarize the plasma membrane and inhibit nerve impulse transmission.

Neurotransmitters and their receptors are targets of pharmacological agents aimed at controlling neurological function. For example, GABA is the major inhibitory neurotransmitter in the CNS, and GABA receptors are the principal target of sedatives such as benzodiazepines and barbiturates which act by enhancing GABA-mediated effects (Katzung, B.G. (1995) Basic and Clinical Pharmacology, 6th edition, Appleton & Lange, Norwalk, CT, pp. 338-339).

Two major classes of neurotransmitter transporters are essential to the function of the nervous system. The first class is uptake carriers in the plasma membrane of neurons and glial cells, which pump neurotransmitters from the extracellular space into the cell. This process relies on the Na^+ gradient across the plasma membrane, particularly the co-transport of Na^+. Two families of proteins have been identified. One family includes the transporters for GABA, monoamines such as noradrenaline, dopamine, and serotonin, and amino acids such as glycine and proline. Common structural components include twelve putative transmembrane a-helical domains, cytoplasmic N- and

C- termini, and a large glycosylated extracellular loop separating transmembrane domains three and four. This family of homologous proteins derives their energy from the co-transport of Na^+ and Cl^- ions with the neurotransmitter into the cell (Na^+/Cl^- neurotransmitter transporters). The second family includes transporters for excitatory amino acids such as glutamate. Common structural components include 6-10 putative transmembrane domains, cytoplasmic N- and C- termini, and glycosylations in the extracellular loops. The excitatory amino acid transporters are not dependent on Cl^-, and may require intracellular K^+ ions (Na^+/K^+- neurotransmitter transporters) (Liu, Y. et al. (1999) Trends Cell Biol. 9:356-363).

The second class of neurotransmitter transporters is present in the vesicle membrane, and concentrates neurotransmitters from the cytoplasm into the vesicle, before exocytosis of the vesicular contents during synaptic transmission. Vesicular transport uses the electrochemical gradient across the vesicular membrane generated by a H^+-ATPase. Two families of proteins are involved in the transport of neurotransmitters into vesicles. One family uses primarily proton exchange to drive transport into secretory vesicles and includes the transporters for monoamines and acetylcholine. For example, the monoamine transporters exchange two luminal protons for each molecule of cytoplasmic transmitter. The second family includes the GABA transporter, which relies on the positive charge inside synaptic vesicles. The two classes of vesicular transporters show no sequence similarity to each other and have structures distinct from those of the plasma membrane carriers (Schloss, P. et al. (1994) Curr. Opin. Cell Biol. 6:595-599; Liu et al., *supra*).

GABA is the predominant inhibitory neurotransmitter and is widely distributed in the mammalian nervous system. GABA is cleared from the synaptic cleft by specific, high-affinity, Na^+- and Cl^-- dependent transporters, which are thought to be localized to both pre- and postsynaptic neurons, as well as to surrounding glial cells. At least four GABA transporters (GAT1–GAT4) have been cloned (Liu, Q.-R. et al. (1993) J. Biol. Chem. 268:2106–2112). Studies of [^{3}H]-GABA uptake into cultured cells and plasma-membrane vesicles isolated from various tissues revealed considerable differences in GABA transporter heterogeneity. GABA transporters exhibit differences in substrate affinity and specificity, distinct blocker pharmacologies, and different tissue localization. For example, the K_m values of GABA uptake of the expressed GAT1 to GAT4 are 6, 79, 18, and 0.8 mM, respectively. In addition to transporting GABA, GAT2 also transports betaine; GAT3 and GAT4 also transport β-alanine and taurine. Pharmacological studies revealed that GABA transport by GAT1 and GAT4 is more sensitive to 2,4-diaminobutyric acid and guavicine than that by GAT2 and GAT3. *In situ* hybridization showed that GAT1 and GAT4 expression is brain specific. GAT2 and GAT3 mRNAs were detected in tissues such as liver and kidney (Schloss et al., *supra*; Borden, L.A. (1996)

Neurochem. Int. 29:335-356; Nelson, N. (1998) J. Neurochem. 71:1785-1803).

Human studies indicated that GABA transporter function is reduced in epileptic hippocampi. Decreased GABAergic neurotransmission has also been implicated in the pathophysiology of schizophrenia (Simpson, M.D. et al. (1992) Psychiatry Res. 42:273-282).

Diazepam binding inhibitor (DBI), also known as endozepine and acyl-Coenzyme (CoA)-binding protein, is an endogenous GABA receptor ligand which is thought to down-regulate the effects of GABA. DBI binds medium- and long-chain acyl-CoA esters with very high affinity and may function as an intracellular carrier of acyl-CoA esters (*125950 Diazepam Binding Inhibitor; DBI, Online Mendelian Inheritance in Man (OMIM); PROSITE PDOC00686 Acyl-CoA-binding protein signature).

Glycine serves as one of the major inhibitory neurotransmitters in the mammalian nervous system by activating chloride-channel receptors, which are members of a ligand-gated ion-channel superfamily (Betz, H. (1990) Neuron 5:383-392). Glycine also facilitates excitatory transmission through an allosteric activation of the N-methyl-D-aspartate (NMDA) receptor (Johnson, J.W. and P. Ascher (1987) Nature 325:529-531). Forms of glycine transporter include GLYT 1 and GLYT 2. Variants of GLYT1 (GLYT1 a/b) are generated by alternative splicing (Liu, Q.-R. et al. (1993) J. Biol. Chem. 268:22802-22808). GLYT1a is transcribed in both neural and non-neural tissues, whereas GLYT1b was detected only in neural tissues (Borowsky, B. et al. (1993) Neuron 10:851-863). High levels of GLYT1a/b mRNA were found in hippocampus and cortex, implying its involvement in the regulation of excitatory synaptic transmission. It is not clear whether GLYT1a is expressed in neurons, in glia or in both. In contrast, GLYT1b is found almost exclusively in fiber tracts, suggesting its localization in glial cells (Schloss et al., *supra*). GLYT2 is expressed mainly in brainstem and spinal cord (Schloss et al., *supra*).

The second identified glycine transporter (GLYT2) differs from GLYT1a/b by its extended intracellular amino terminus. The predominant localization of its mRNA in brainstem and spinal cord and its insensitivity to N-methyl-aminoacetic acid suggests that GLYT2 terminates signal transduction at the strychnine-sensitive inhibitory glycine receptor. It has been proposed that, upon depolarization of cells harboring GLYT1b, the transporter runs backwards and releases glycine to act as a neuromodulatory amino acid at the NMDA receptor (Attwell, D. and M. Bouvier (1992) Curr. Biol. 2:541-543). Such a Ca^{2+}-independent, non-vesicular release of neurotransmitters by reverse transport was demonstrated for glutamate and serotonin. This evidence suggests that the transmitter transporters may be important for both the initiation and termination of neurotransmitter action (Schloss et al., *supra*).

The plasma membrane dopamine transporter (DAT) is essential for the reuptake of released dopamine from the synapse. Uptake of dopamine is temperature- and time-dependent, and is inhibited by a variety of compounds, such as cocaine. DAT- knockout mice have been shown to exhibit extreme hyperactivity and resistance to both cocaine and amphetamine, consistent with the primary action of cocaine on DAT (Giros, B. et al. (1996) Nature 379:606-612). The perturbation of the tightly regulated DAT also predisposes neurons to damage by a variety of insults. Most notable is the selective degeneration of DAT-expressing dopamine nerve terminals in the striatum thought to underlie Parkinson's disease. DAT expression can predict the selective vulnerability of neuronal populations, which suggests that therapeutic strategies aimed at altering DAT function could have significant benefits in a variety of disorders (Gary, W.M. et al. (1999) Trends Pharmacol. Sci. 20:424-429).

Creatine transporters are strongly related to transporters for GABA. The primary sequence identity between creatine transporter species homologs is very high (98-99%). Pharmacological characterization demonstrated high affinity creatine uptake (27-43 mM), which was blocked by creatine analogs with high affinity. Creatine transporters are widely expressed in a variety of mammalian tissues, including brain, adrenal gland, intestine, colon, prostate, thymus, ovary, spleen, pancreas, placenta, umbilical cord, thyroid, tongue, pharynx, vertebral discs, jaw, and nasal epithelium. Genetic mapping in the mouse localizes the creatine transporter to a region on the X chromosome in linkage conservation with the human region Xq28, the location of the genes for several neuromuscular diseases (Nash, S.R. et al. (1994) Receptors Channels 2:165-174).

The substrates of a number of cDNA clones encoding proteins of the Na^+ /Cl^--dependent transporter families are still not identified. These are orphan transporters. Identification of the substrates for orphan transporters has been difficult because *in situ* hybridization and immunohistochemistry indicate that the transporters are synthesized by phenotypically different neuronal populations, for example glutaminergic, GABAergic, histaminergic, or serotoninergic neurons. One of the transporters, NTT4, exhibits the highest homology to the creatine transporter. It differs structurally from other members of this family in having an unusually long loop between transmembranes seven and eight (Liu, Q.-R. et al. (1993) FEBS Lett. 315:114-118; Schloss et al., *supra*).

Glutamate is a major excitatory neurotransmitter in the mammalian central nervous system. Electrogenic (Na^+ /K^+)-coupled glutamate transporters, located in the plasma membranes of nerve terminals and glial cells, mediate removal of glutamate released at excitatory synapses and maintain extracellular concentrations below neurototoxic levels. Glutamate transporters achieve this process by

co-transport with three sodium ions and one proton, followed by translocation of a potassium ion in the opposite direction (Zerangue, N. and M.P. Kavanaugh (1996) Nature 383:634-637).

Glutamate transporters belong to a large family of transport proteins. The membrane topology of the glutamate transporters reveals six membrane-spanning helices in the N-terminal part of the proteins (Slotboom, D.J. et al. (1999) Microbiol. Mol. Biol. Rev. 63:293-307). The C-terminal half of the glutamate transporters is well conserved and constitutes a major part of the translocation pathway and contains the binding sites for the substrate and co-transported ions (Zhang, Y. and B.I. Kanner (1999) Proc. Natl. Acad. Sci. USA 96:1710-1715).

The membrane topology of the glutamate transporters reveals six membrane-spanning helices in the N-terminal part of the proteins (Slotboom, D.J. et al. (1999) Microbiol. Mol. Biol. Rev. 63:293-307). The C-terminal half of the glutamate transporters is well conserved and constitutes a major part of the translocation pathway and contains the binding sites for the substrate and co-transported ions (Zhang, Y. and B.I. Kanner (1999) Proc. Natl. Acad. Sci. USA 96:1710-1715).

Impaired re-uptake of synaptic glutamate, and a reduced expression of glutamate transporters have been found in the motor cortex of patients with amyotrophic lateral sclerosis (ALS). Inhibition of the synthesis of each glutamate transporter subtype using chronic antisense oligonucleotide administration, *in vitro* and *in vivo*, selectively and specifically reduced the protein expression and function of glutamate transporters. The loss of glial glutamate transporters produced elevated extracellular glutamate levels, neurodegeneration characteristic of excitotoxicity, and a progressive paralysis. The loss of the neuronal glutamate transporter did not elevate extracellular glutamate in the striatum but produced mild neurotoxicity and resulted in epilepsy (Rothstein, J.D. et al. (1996) Neuron 16:675-686).

The vesicular monoamine transporters (VMAT) package cytoplasmic monoamine neurotransmitters into secretory vesicles for regulated exocytotic release. VMAT acts as an electrogenic exchanger of protons and monoamines, using a proton electrochemical gradient. VMAT transporters include VMAT1 and VMAT2. The VMAT proteins possess twelve transmembrane segments, with both extremities lying on the cytoplasmic side. VMAT proteins are associated with distinct vesicle populations in neurons and neuroendocrine cells (Henry, J.-P. et al. (1994) J. Exp. Biol. 196:251-262).

Vesicular transport is inhibited by the antihypertensive drug reserpine and the related but more centrally acting drug tetrabenazine. The mechanism of transport and the biochemistry of VMAT have been analyzed with these drugs, using mainly the chromaffin granules from bovine adrenal glands as a source of transporters (Peter, D. et al. (1994) J. Biol. Chem. 269:7231-7237).

Human studies indicated that reserpine can cause a syndrome resembling depression, indicating the importance of vesicular transport activity for the control of mood and behavior. The psychostimulant amphetamine also disrupts the storage of amines in secretory vesicles, further indicating that alterations in vesicular monoamine transport can affect behavior (Sulzer, D. and S. Rayport (1990) Neuron 5:797-808).

Human diseases caused by defects in neurotransmitter transporters include schizophrenia, Tourette's syndrome, Parkinson's disease, brain ischemia, amyotrophic lateral scerlosis, depression, and epilepsy. For example, decreased GABAergic neurotransmission has been implicated in the pathophysiology of CNS disorders such as epilepsy and schizophrenia. Impaired re-uptake of synaptic glutamate, and a reduced expression of the glutamate transporter have been found in the motor cortex of patients with amyotrophic lateral sclerosis (ALS). The loss of glial glutamate transporters produces elevated extracellular glutamate levels, neurodegeneration characteristic of excitotoxicity, and a progressive paralysis. The loss of neuronal glutamate transporters produces mild neurotoxicity and result in epilepsy (Rothstein, J.D. et al. (1996) Neuron 16:675-686).

Transporters for dopamine, norepinephrine, and serotonin have particular significance as targets for clinically relevant psychoactive agents including cocaine, antidepressants, and amphetamines. Cocaine and antidepressants are transporter antagonists that act with varying degrees of specificity to enhance synaptic concentrations of amines by limiting clearance. Amphetamines enhance transporter mediated efflux in concert with a depletion of vesicular amine stores (Barker, E.L. and R.D. Blakely (1995) Psychopharmacology 28:321-333; Sulzer, D. and S. Rayport (1990) Neuron 5:797-808; Wall, S.C. et al. (1995) Mol. Pharmacol. 47:544-550).

Each of over a trillion neurons in adult humans connects with over a thousand target cells (Tessier-Lavigne, M. et al. (1996) Science 274:1123-1133). These neuronal connections form during embryonic development. Each differentiating neuron sends out an axon tipped at the leading edge by a growth cone. Aided by molecular guidance cues, the growth cone migrates through the embryonic environment to its synaptic target. Semaphorins are growth cone guidance signals that may function during embryogenesis by providing local signals to specify territories inaccessible to growing axons (Puschel, A.W. et al. (1995) Neuron 14:941-948).

The guidance of axons during development involves both positive and negative effects (*i.e.*, chemoattraction and chemorepulsion). The Slit family of proteins have been implicated in promoting axon branching, elongation, and repulsion. Members of the Slit family have been identified in a variety of organisms, including insects, amphibians, birds, rodents and humans (Guthrie, S. (1999) Current Biology 9:R432-R435). Slit proteins appear to be ligands for the repulsive guidance receptor,

Roundabout (Robo); however, Slit proteins also cause elongation in some assays. A post-translationally processed form of Slit appears to be the active form of the protein (Guthrie, S. supra and Brose, K. et al. (1999) Cell 96:795-806).

Axon growth is also guided in part by contact-mediated mechanisms involving cell surface and extracellular matrix (ECM) molecules. Many ECM molecules, including fibronectin, vitronectin, members of the laminin, tenascin, collagen, and thrombospondin families, and a variety of proteoglycans, can act either as promoters or inhibitors of neurite outgrowth and extension (Tessier-Lavigne et al., supra). Receptors for ECM molecules include integrins, immunoglobulin superfamily members, and proteoglycans. ECM molecules and their receptors have also been implicated in the adhesion, maintenance, and differentiation of neurons (Reichardt, L.F. et al. (1991) Ann. Rev. Neurosci. 14:531-571). The proteoglycan testican is localized to the post-synaptic area of pyramidal cells of the hippocampus and may play roles in receptor activity, neuromodulation, synaptic plasticity, and neurotransmission (Bonnet, F. et al. (1996) J. Biol. Chem. 271:4373-4380).

Other nervous system-associated proteins have roles in neuron signaling, cell adhesion, nerve regeneration, axon guidance, and neurogenesis. The neurexophilins are neuropeptide-like proteins which are proteolytically processed after synthesis. They are ligands for the neuron-specific cell surface proteins, the α-neurexins. Neurexophilins and neurexins may participate in a neuron signaling pathway (Missler, M. and T.C. Sudhof (1998) J. Neurosci. 18:3630-3638; Missler, M. et al. (1998) J. Biol. Chem. 273:34716-34723). Ninjurin is a neuron cell surface protein which plays a role in cell adhesion and in nerve regeneration following injury. Ninjurin is up-regulated after nerve injury in dorsal root ganglion neurons and in Schwann cells (*602062 Ninjurin; NINJ1 OMIM; Araki, T. and Milbrandt, J. (1996) Neuron 17:353-361). Mammalian Numb is a phosphotyrosine-binding (PTB) domain-containing protein which may be involved in cortical neurogenesis and cell fate decisions in the mammalian nervous system. Numb's binding partner, the LNX protein, contains four PDZ domains and a ring finger domain and may participate in a signaling pathway involving Numb. PDZ domains have been found in proteins which act as adaptors in the assembly of multifunctional protein complexes involved in signaling events at surfaces of cell membranes (Ponting, C.P. (1997) Bioessays 19:469-479). LNX contains a tyrosine phosphorylation site which may be important for the binding of other PTB-containing proteins such as SHC, an adaptor protein which associates with tyrosine-phosphorylated growth factor receptors and downstream effectors (Dho, S.E. et al. (1998) J. Biol. Chem. 273:9179-9187).

Another family of molecules that appear to be important for neurotransmission are the choline-transporter-like CTL1 proteins. The prototypic CTL1 was identified in yeast as a suppressor

of a choline transport mutation; however, mammalian homologues have been identified. The proteins comprise approximately ten putative transmembrane domains in addition to transporter-like motifs but do not appear to be canonical choline transporters. Choline transport is important to neurotransmission because choline is a precursor of acetylcholine, required in abundance by cholinergic neurons (O'Regan, S. et al. (2000) Proc. Natl. Acad. Sci. U.S.A. 97:1835-1840).

Transcriptional regulatory proteins are also essential for the development of the nervous system and elements of neurotransmission. A specific class of transcription factors, homeobox transcription factors, directs nerve cell-associated tissue patterning and differentiation. The presence and function of these proteins appears to be ubiquitous in nematodes, arthropods, and vertebrates. One example of these proteins is DRG11, a homeobox transcription factor expressed in mammalian sensory neurons, and which appears to be involved in neural crest development (Saito, T. et al. (1995) Mol. Cell Neurosci. 6:280-292).

cDNA clones were selected by comparing libraries of normal mouse cerebellar cDNA and cerebellar cDNA from Purkinje cell degeneration (pcd) mice. An mRNA present in Purkinje neurons encodes PCD5, a protein of 99 amino acids. PCD5's expression is restricted to the cerebellum and the eye. The gene encoding PCD5 was localized to mouse chromosome 8 (Nordquist, D.T. et al. (1988) J. Neurosci. 8(12):4780-4789).

Neuronal signals are transmitted across the neuromuscular junction (NMJ). Motor axons release the molecule agrin to induce the formation of the postsynaptic apparatus in muscle fibers. Proteins such as dystroglycan, MuSK, and rapsyn participate in the transduction of agrin signals. Agrin also functions in the upregulation of gene transcription in myonuclei and the control of presynaptic differentiation (Ruegg, M.A. and J.L. Bixby (1998) Trends Neurosci. 21:22-27).

Neurological protein domains

CNS-associated proteins can be phosphoproteins. For example, ARPP-21 (cyclic AMP-regulated phosphoprotein) is a cytosolic neuronal phosphoprotein that is highly enriched in the striatum and in other dopaminoceptive regions of the brain. The steady-state level of ARPP-21 mRNA is developmentally regulated. But, in the neonatal and mature animal, ARPP-21 mRNA is not altered following 6-hydroxydopamine lesions of the substantia nigra or by pharmacologic treatments that upregulate the D1- or D2-dopamine receptors (Ehrlich, M.E. et al. (1991) Neurochem. 57:1985-1991).

CNS-associated signaling proteins may contain PDZ domains. PDZ domains have been found in proteins which act as adaptors in the assembly of multifunctional protein complexes involved in signaling events at surfaces of cell membranes. PDZ domains are generally found in membrane-

associated proteins including neuronal nitric oxide synthase (NOS) and several dystrophin-associated proteins (Ponting, C.P. et al. (1997) Bioessays 19:469-479). PSD-95/SAP90 is a membrane-associated guanylate kinase found in neuronal cells at the postsynaptic density (PSD) (Takeuchi, M. et al. (1997) J. Biol. Chem. 272:11943-11951). PSD-95/SAP90 contains three PDZ domains, one SH3 domain, and one guanylate kinase domain. The PDZ domains mediate interactions with NMDA receptors, Shaker-type potassium channels, and brain nitric oxide synthase. SAPAPs (SAP90/PSD-95-Associated Proteins) promote localization of PSD-95/SAP90 at the plasma membrane.

CNS-associated proteins may also contain epidermal growth factor (EGF) domains. The Notch proteins are transmembrane proteins which contain extracellular regions of repeated EGF domains. Notch proteins, such as the *Drosophila melanogaster* neurogenic protein Notch, are generally involved in the inhibition of developmental processes. Other members of the Notch family are the lin-12 and glp-1 genes of *Caenorhabditis elegans*. Genetic studies indicate that the lin-12 and glp-1 proteins act as receptors in specific developmental cell interactions which may be involved in certain embyronic defects (Tax, F. E. et al. (1994) Nature 368:150-154). Pecanex, a maternal-effect neurogenic locus of *D. melanogaster*, is believed to encode a large transmembrane protein. In the absence of maternal expression of the pecanex gene, an embryo develops severe hyperneuralization similar to that characteristic of Notch mutant embryos (LaBonne, S. G. et al. (1989) Dev. Biol. 136:1-116).

Other CNS-associated signaling proteins contain WW domains. The WW domain is a protein motif with two highly conserved tryptophans. It is present in a number of signaling and regulatory proteins, including Huntingtin interacting protein. Several fibroblast growth factor (FGF) homologous factors (i.e., FHF polypeptides) have also been implicated in nervous system development based on mRNA expression patterns in mouse and human tissues. Members of the FHF family of polypeptides are structurally distinct from prototypic FGFs, consistent with the unusual role of these FGF-related proteins (Smallwood, P.M. et al. (1996) Proc. Natl. Acad. Sci. U.S.A. 93:9850-9857 and Hartung, H. et al. (1997) Mech. Dev. 64:31-39).

CNS-associated proteins may also contain leucine rich repeats (LRR) which are short motifs found in numerous proteins from a wide range of species. LRR motifs are of variable length, most commonly 20-29 amino acids, and multiple repeats are typically present in tandem. LRR motifs are important for protein/protein interactions, and LRR proteins are involved in cell/cell interactions, morphogenesis, and development (Kobe, B. and Deisenhofer, J. (1995) Curr. Opin. Struct. Biol. 5:409-416). The human ISLR (immunoglobulin superfamily containing leucine-rich repeat) protein contains a C2-type immunoglobulin domain as well as LRR motifs. The ISLR gene is linked to the critical region

for Bardet-Biedl syndrome, a developmental disorder of which the most common feature is retinal dystrophy (Nagasawa, A. et al. (1999) Genomics 61:37-43). Brain leucine rich repeat protein, a member of the LRR family of proteins, possibly functions in nervous system development and maintenance (Taniguchi, H. et al. (1996) Brain Res. Mol. Brain Res. 36:45-52).

Disorders associated with neurological processes

Alzheimer's disease (AD) is a degenerative disorder of the CNS which causes progressive memory loss and cognitive decline during mid to late adult life. AD is characterized by a wide range of neuropathologic features including amyloid deposits and intra-neuronal neurofibrillary tangles. Although the pathogenic pathway leading to neurodegeneration and AD is not well understood, at least three genetic loci that confer genetic susceptibility to the disease have been identified (Schellenberg, G.D. (1995) Proc. Natl. Acad. Sci. 92:8552-8559; Sherrington, R. et al. (1995) Nature 375:754-760).

Familial British dementia (FBD), is an autosomal dominant disease featuring amyloid plaques surrounded by astrocytes and microglia, neurofibrillary tangles, neuronal loss, and progressive dementia. The BRI gene on chromosome 13 encodes a 4 kD peptide, A-Bri. This membrane-anchored protein is a primary constituent of amyloid deposits, and its presence in lesions from the CNS of FBD patients may be a contributive factor of this disease (El-Agnaf, O.M.A. et al. (2001) Biochemistry 40:3449-3457).

Astrocytomas, and the more malignant glioblastomas, are the most common primary tumors of the brain, accounting for over 65% of primary brain tumors. These tumors arise in glial cells of the astrocyte lineage. Following infection by pathogens, astrocytes function as antigen-presenting cells and modulate the activity of lymphocytes and macrophages. Astrocytomas constitutively express many cytokines and interleukins that are normally produced only after infection by a pathogen (de Micco, C. (1989) J. Neuroimmunol. 25:93-108). In the course of identifying genes related to astrocyte differentiation, one cDNA was isolated from an astrocytoma cDNA library that encodes a protein structurally related to the plant pathogenesis-related (PR) proteins (Murphy, E.V. et al. (1995) Gene 159:131-135). The glioma pathogenesis-related protein (GliPR) is highly expressed in glioblastoma, but not in fetal or adult brain, or in other nervous system tumors. PR proteins are a family of small (10-20 kDa), protease resistant proteins induced in plants by viral infections, such as tobacco mosaic virus. The synthesis of PR proteins is believed to be part of a primitive immunological response in plants (van Loon, L.C. (1985) Plant Mol. Biol. 4:111-116). GliPR shares up to 50% homology with the PR-1 protein family over a region that comprises almost two thirds of the protein, including a conserved triad of amino acids, His-Glu-His, appropriately spaced to form a metal-binding domain (Murphy et al., *supra*).

Signaling initiated by the Trk family receptors plays a dynamic role in neurogenic tumors. The proto-oncogene Trks encode the high-affinity receptor tyrosine kinases for nerve growth factor (NGF) neurotrophins. A rearranged Trk oncogene is often observed in non-neuronal neoplasms such as colon and papillary thyroid cancers. The proto-oncogene Trks regulates growth, differentiation and apoptosis of tumors of neuronal origin, such as neuroblastoma and medulloblastoma (Nakagawara, A. (2001) Cancer Lett.169:107-114).

Neuronal thread proteins (NTP) are a group of immunologically related molecules found in the brain and neuroectodermal tumor cell lines. NTP expression is increased in neuronal cells during proliferation, differentiation, brain development, in Alzheimer's disease (AD) brains, and in pathological states associated with regenerative nerve sprouting (de la Monte, S.M. et al. (1996) J. Neuropathol. Exp. Neurol. 55:1038-1050). Monoclonal antibodies generated to a recombinant NTP, AD7c-NTP, isolated from an end-stage AD brain library, showed high levels of NTP immunoreactivity in perikarya, neuropil fibers, and white matter fibers of AD brain tissue. *In vitro* studies also demonstrated NTP upregulation, phosphorylation, and translocation from the perikarya to cell processes and growth cones during growth factor-induced neuritic sprouting and neuronal differentiation. Additionally, increased NTP immunoreactivity was found in Down syndrome brains beginning in the second decade, prior to establishment of widespread AD neurodegeneration, and at an age when a low-level or an absence of NTP expression was observed in control brains. These findings indicated that abnormal expression and accumulation of NTP in brain may be an early marker of AD neurodegeneration in Down syndrome (de la Monte, S.M. et al. (1996) J. Neurol. Sci. 135:118-125). Furthermore, the increased expression and accumulation of NTP in AD brain tissue was paralleled by corresponding elevations of NTP in cerebrospinal fluid (CSF), and elevated levels of NTP were detectable in the CSF early in the course of the disease.

Ubiquilin is a presenilin-interacting protein. Ubiquilin contains numerous ubiquitin-like domains thought to be involved in targeting proteins for degradation. However, ubiquilin promotes increased presenilin protein accumulation. Ubiquilin, as a modulator of presenilin levels may have implications involving various cellular functions. Presenilins are linked to a variety of biological processes, including calcium regulation, Notch signaling, apoptosis, regulation of the cell cycle, including the unfolded-protein response and β-amyloid precursor protein-associated gamma secretase activity. Expression of ubiquilin is highest in human brain neurons and is associated with neurofibrillary tangles and Lewy bodies of Alzheimer's disease and Parkinson's disease brains, respectively (Mah,A.L., et al. (2000) J. Cell Biol. 151:847-862).

Fe65-like protein (Fe65L2), a new member of the Fe65 protein family, is one of the ligands

that interacts with the cytoplasmic domain of Alzheimer beta-amyloid precursor protein (APP). Transgenic mice expressing APP simulate some of the prominent behavioral and pathological features of Alzheimer's disease, including age-related impairment in learning and memory, neuronal loss, gliosis, neuritic changes, amyloid deposition, and abnormal tau phosphorylation (Duilio, A. et al. (1998) Biochem. J. 330:513-519).

Amyotrophic lateral sclerosis (ALS) is characterized by motor neuron death, altered peroxidase activity of mutant SOD1, changes in intracellular copper homeostasis, protein aggregation, and changes in the function of glutamate transporters leading to excitotoxicity. Neurofilaments and peripherin appear to play some part in motor neuron degeneration. ALS is occasionally associated with mutations of the neurofilament heavy chain gene (Al-Chalabi, A. and P.N. Leigh (2000) Curr. Opin. Neurol. 13:397-405). Cytoskeletal abnormalities such as abnormal inclusions containing neurofilaments (NFs) and/or peripherin, reduced mRNA levels for the NF light (NF-L) protein and mutations in the NF heavy (NF-H) gene have been observed in ALS. Intermediate filament inclusions containing peripherin may play a contributory role in ALS (Julien, J.P. and J.M. Beaulieu (2000) J. Neurol. Sci.180:7-14).

Miller-Dieker syndrome (MDS) or isolated lissencephaly syndrome (ILS) are characterized by a smooth cerebral surface, a thickened cortex with four abnormal layers, and misplaced neurons. Both conditions may result from deletion or mutation in the LIS1 gene. The lissencephaly gene product Lis1 is a component of evolutionarily conserved intracellular multiprotein complexes essential for neuronal migration, and which may be components of the machinery for cell proliferation and intracellular transport (Leventer, R.J. et al. (2001) Trends Neurosci. 24:489-492). NudC, a nuclear movement protein, interacts with Lis1 (Morris, S.M. et al. (1998) Curr. Biol. 8:603-606).

CNS-associated proteins can also be phosphoproteins. For example, ARPP-21 (cyclic AMP-regulated phosphoprotein) is a cytosolic neuronal phosphoprotein that is highly enriched in the striatum and in other dopaminoceptive regions of the brain. The steady-state level of ARPP-21 mRNA is developmentally regulated. But, in the neonatal and mature animal, ARPP-21 mRNA is not altered following 6-hydroxydopamine lesions of the substantia nigra or by pharmacologic treatments that upregulate the D1- or D2-dopamine receptors. (Ehrlich, M. E. et al. (1991) Neurochem. 57:1985-1991.)

Retinitis pigmentosa comprises a group of slowly progressive, inherited disorders of the retina that cause loss of night vision and peripheral visual field in adolescence. A recessive nonsense mutation in the *Drosophila* opsin gene causes photoreceptor degeneration. In some families, genes encoding rhodopsin and peripherin/RDS map very close to the disease loci. Rhodopsin and

peripherin/RDS mutations have been found in approximately 30% of all autosomal dominant cases (Shastry, B.S. (1994) Am. J. Med. Genet. 52:467-474).

Astrocytomas, and the more malignant glioblastomas are primary tumors of the brain. The glioma pathogenesis-related protein (GliPR) is highly expressed in glioblastoma, but not in fetal or adult brain, or in other nervous system tumors (Murphy, E.V. et al. (1995) Gene 159:131-135). Signaling initiated by the Trk family receptors plays a dynamic role in neurogenic tumors. The proto-oncogene Trks encode the high-affinity receptor tyrosine kinases for nerve growth factor (NGF) neurotrophins. A rearranged Trk oncogene is often observed in non-neuronal neoplasms such as colon and papillary thyroid cancers. The proto-oncogene Trks regulates growth, differentiation and apoptosis of tumors of neuronal origin, such as neuroblastoma and medulloblastoma (Nakagawara, A. (2001) Cancer Lett.169:107-114).

Synaptic proteins are involved in Alzheimer's disease (AD) and other disorders including ischemia, a variety of disorders where synapse-associated proteins are abnormally accumulated in the nerve terminals or synaptic proteins are altered after denervation, and neoplastic disorders (Masliah, E. and R. Terry (1993) Brain Pathol. 3:77-85). Synaptophysin (SY), a major integral membrane protein of small synaptic vesicles, is on the X chromosome in subbands Xp11.22-p11.23, a region implicated in several inherited diseases including Wiskott-Aldrich syndrome, three forms of X-linked hypercalciuric nephrolithiaisis, and the eye disorders retinitis pigmentosa 2, congenital stationary night blindness, and Aland Island eye disease (Fisher, S.E. et al. (1997) Genomics 45:340-347).

Mutations in the BRI2 isoform of the BRI gene family are associated with dementia in humans (Vidal, R. et al. (2001) Gene 266:95-102).

Changes in the molecular and cellular components of neuronal signaling systems correlate with the effects on mood and cognition observed after long-term treatment with antidepressant drugs. Two serine/threonine kinases, Ca^{2+}/calmodulin-dependent protein kinase II and cyclic AMP-dependent protein kinase, are activated in the brain following antidepressant treatment. Associated changes in the phosphorylation of selected protein substrates in subcellular compartments including presynaptic terminals and microtubules may contribute to the modulation of synaptic transmission observed with antidepressants (Popoli, M. et al. (2001) Pharmacol. Ther. 89:149-170). Reserpine can cause a syndrome resembling depression, indicating the importance of vesicular transport activity for the control of mood and behavior. The psychostimulant amphetamine also disrupts the storage of amines in secretory vesicles, further indicating that alterations in vesicular monoamine transport can affect behavior (Sulzer, D. and S. Rayport (1990) Neuron 5:797-808).

Decreased GABAergic neurotransmission has been implicated in the pathophysiology of CNS

disorders such as epilepsy and schizophrenia. Impaired re-uptake of synaptic glutamate and a reduced expression of the glutamate transporter have been found in the motor cortex of patients with amyotrophic lateral sclerosis (ALS). The loss of glial glutamate transporters produces elevated extracellular glutamate levels, neurodegeneration characteristic of excitotoxicity, and a progressive paralysis. The loss of neuronal glutamate transporters produces mild neurotoxicity and results in epilepsy (Rothstein, J.D. et al. (1996) Neuron 16:675-686). GABA transporter function is reduced in epileptic hippocampi. Transporters for dopamine, norepinephrine, and serotonin have particular significance as targets for clinically relevant psychoactive agents including cocaine, antidepressants, and amphetamines. Cocaine and antidepressants are transporter antagonists that act with varying degrees of specificity to enhance synaptic concentrations of amines by limiting clearance. Amphetamines enhance transporter mediated efflux in concert with a depletion of vesicular amine stores (Barker, E.L. and R.D. Blakely (1995) Psychopharmacology 28:321-333; Sulzer, D. and S. Rayport (1990) Neuron 5:797-808; Wall, S.C. et al. (1995) Mol. Pharmacol. 47:544-550).

The μ-opioid receptor (MOR) mediates the actions of analgesic agents including morphine, codeine, methadone, and fentanyl as well as heroin. MOR is functionally coupled to a G-protein-activated potassium channel (Mestek A. et al. (1995) J. Neurosci. 15:2396-2406). A variety of MOR subtypes exist. Alternative splicing has been observed with MOR-1 as with a number of G protein-coupled receptors including somatostatin 2, dopamine D2, prostaglandin EP3, and serotonin receptor subtypes 5-hydroxytryptamine4 and 5-hydroxytryptamine7 (Pan, Y.X. et al. (1999) Mol. Pharm. 56:396-403).

The central nervous system regulates the innate immune system by elaborating anti-inflammatory hormone cascades in response to bacterial products and immune mediators. The central nervous system also responds via acetylcholine-mediated efferent signals carried through the vagus nerve. Nicotinic cholinergic receptors expressed on macrophages detect these signals and respond with a dampened cytokine response (Tracey K.J. et al. (2001) FASEB J.15:1575-1576).

Machado-Joseph disease (MJD) is an autosomal dominant, neurodegenerative disorder characterized by cerebellar ataxia, pyramidal and extra-pyramidal signs, peripheral nerve palsy, external ophthalmoplegia, facial and lingual fasciculation and bulging. The MJD josephin protein (amino acid residues 29-62) is a predominantly cytoplasmic protein associated with human neurons, but also detected in the nuclei of neurons and glial cells (Mruk, D.D. and Cheng, C.Y. (1999) J. Biol. Chem. 274:27056-27068).

Juvenile neuronal ceroid lipofuscinosis (JNCL), also known as Batten disease, is an autosomal

recessive lysosomal storage disease associated with mutations in CLN3. The predominant mutation in CLN3 is a 1.02 kb genomic deletion that accounts for nearly 85% of the disease alleles. Additional missense and nonsense mutations have been described. Some missense substitutions result in a protracted phenotype, with delays in the onset of classical clinical features, whereas others lead to classical JNCL. CLN3 is a hydrophobic protein containing 5 to 7 transmembrane domains. CLN3 is found to be highly associated with lysosome-associated membrane protein II in non-neuronal cells and with synaptophysin in neuronal cell lines (Haskell, R.E. et al. (2000) Hum. Mol. Genet. 9:735-744).

Dysferlin is the protein product of the gene mutated in patients with an autosomal recessive limb-girdle muscular dystrophy type 2B (LGMD2B) and a distal muscular dystrophy, Miyoshi myopathy. Dysferlin is homologous to a *Caenorhabditis elegans* spermatogenesis factor, FER-1. Otoferlin, another human FER-1-like protein (ferlin), is responsible for autosomal recessive nonsyndromic deafness (DFNB9). All the ferlins are characterized by sequences corresponding to multiple C2 domains that share the highest level of homology with the C2A domain of rat synaptotagmin III (Britton S. et al. (2000) Genomics 68:313-321).

Expression profiling

Microarrays are analytical tools used in bioanalysis. A microarray has a plurality of molecules spatially distributed over, and stably associated with, the surface of a solid support. Microarrays of polypeptides, polynucleotides, and/or antibodies have been developed and find use in a variety of applications, such as gene sequencing, monitoring gene expression, gene mapping, bacterial identification, drug discovery, and combinatorial chemistry.

One area in particular in which microarrays find use is in gene expression analysis. Array technology can provide a simple way to explore the expression of a single polymorphic gene or the expression profile of a large number of related or unrelated genes. When the expression of a single gene is examined, arrays are employed to detect the expression of a specific gene or its variants. When an expression profile is examined, arrays provide a platform for identifying genes that are tissue specific, are affected by a substance being tested in a toxicology assay, are part of a signaling cascade, carry out housekeeping functions, or are specifically related to a particular genetic predisposition, condition, disease, or disorder.

Lung Cancer

Lung cancer is the leading cause of cancer death for men and the second leading cause of cancer death for women in the U.S. The vast majority of lung cancer cases are attributed to smoking tobacco, and increased use of tobacco products in third world countries is projected to lead to an

epidemic of lung cancer in these countries. Exposure of the bronchial epithelium to tobacco smoke appears to result in changes in tissue morphology, which are thought to be precursors of cancer. Lung cancers are divided into four histopathologically distinct groups. Three groups (squamous cell carcinoma, adenocarcinoma, and large cell carcinoma) are classified as non-small cell lung cancers (NSCLCs). The fourth group of cancers is referred to as small cell lung cancer (SCLC). Collectively, NSCLCs account for ~70% of cases while SCLCs account for ~18% of cases. The molecular and cellular biology underlying the development and progression of lung cancer are incompletely understood.

Deletions on chromosome 3 are common in this disease and are thought to indicate the presence of a tumor suppressor gene in this region. Activating mutations in K-ras are commonly found in lung cancer and are the basis of one of the mouse models for the disease.

Ovarian Cancer

Ovarian cancer is the leading cause of death from a gynecologic cancer. The majority of ovarian cancers are derived from epithelial cells, and 70% of patients with epithelial ovarian cancers present with late-stage disease. As a result, the long-term survival rates for this disease is very low. Identification of early-stage markers for ovarian cancer would significantly increase the survival rate. The molecular events that lead to ovarian cancer are poorly understood. Some of the known aberrations include mutation of p53 and microsatellite instability. Since gene expression patterns are likely to vary when normal ovary is compared to ovarian tumors, examination of gene expression in these tissues to identify possible markers for ovarian cancer is particularly relevant to improving diagnosis, prognosis, and treatment of this disease.

Colon Cancer

Colorectal cancer is the second leading cause of cancer deaths in the United States. Colon cancer is associated with aging, since 90% of the total cases occur in individuals over the age of 55. A widely accepted hypothesis is that several contributing genetic mutations must accumulate over time in an individual who develops the disease. To understand the nature of genetic alterations in colorectal cancer, a number of studies have focused on the inherited syndromes. The first known inherited syndrome, Familial Adenomatous Polyposis (FAP), is caused by mutations in the Adenomatous Polyposis Coli gene (APC), resulting in truncated or inactive forms of the protein. This tumor suppressor gene has been mapped to chromosome 5q. The second known inherited syndrome is hereditary nonpolyposis colorectal cancer (HNPCC), which is caused by mutations in mismatch repair genes.

Although hereditary colon cancer syndromes occur in a small percentage of the population and most colorectal cancers are considered sporadic, knowledge from studies of the hereditary syndromes can be generally applied. For instance, somatic mutations in APC occur in at least 80% of indiscriminate colon tumors. APC mutations are thought to be the initiating event in the disease. Other mutations occur subsequently. Approximately 50% of colorectal cancers contain activating mutations in ras, while 85% contain inactivating mutations in p53. Changes in these genes lead to gene expression changes in colon cancer. Less is understood about downstream targets of these mutations and the role they may play in cancer development and progression.

Tangier Disease

Tangier disease (TD) is a rare genetic disorder characterized by near absence of circulating high density lipoprotein (HDL) and the accumulation of cholesterol esters in many tissues, including tonsils, lymph nodes, liver, spleen, thymus, and intestine. Low levels of HDL represent a clear predictor of premature coronary artery disease and homozygous TD correlates with a four- to six-fold increase in cardiovascular disease compared to controls. The major cardio-protective activity of HDL is ascribed to its role in reverse cholesterol transport, the flux of cholesterol from peripheral cells such as tissue macrophages, through plasma lipoproteins to the liver. The HDL protein, apolipoprotein A-I, plays a major role in this process, interacting with the cell surface to remove excess cholesterol and phospholipids. Recent studies have shown that this pathway is severely impaired in TD and the defect lies in a specific gene, the ABC1 transporter. This gene is a member of the family of ATP-binding cassette transporters, which utilize ATP hydrolysis to transport a variety of substrates across membranes.

RNA Expression

Atherosclerosis and the associated coronary artery disease and cerebral stroke represent the most common cause of death in industrialized nations. Although certain key risk factors have been identified, a full molecular characterization that elucidates the causes and provide care for this complex disease has not been achieved. Molecular characterization of growth and regression of atherosclerotic vascular lesions requires identification of the genes that contribute to features of the lesion including growth, stability, dissolution, rupture and, most lethally, induction of occlusive vessel thrombus.

An early step in the development of atherosclerosis is formation of the "fatty streak". Lipoproteins, such as the cholesterol-rich low-density lipoprotein (LDL), accumulate in the extracellular space of the vascular intima, and undergo modification. Oxidation of LDL occurs most

avidly in the sub-endothelial space where circulating antioxidant defenses are less effective. The degree of LDL oxidation affects its interaction with target cells. "Minimally oxidized" LDL (MM-LDL) is able to bind to LDL receptor but not to the oxidized LDL (Ox-LDL) or "scavenger" receptors that have been identified, including scavenger receptor types A and B, CD36 , CD68/macrosialin and LOX-1 (Navab et al. (1994) Arterioscler Thromb Vasc Biol 16:831-842; Kodama et al. (1990) Nature 343:531-535; Acton et al. (1994) J Biol Chem 269:21003-21009; Endemann et al. (1993) J Biol Chem 268:11811-11816; Ramprasad et al. (1996) Proc Natl Acad Sci 92:14833-14838; Kataoka *et al.* (1999) Circulation 99:3110-3117). MM-LDL can increase the adherence and penetration of monocytes, stimulate the release of monocyte chemotactic protein 1 (MCP-1) by endothelial cells, and induce scavenger receptor A (SRA) and CD36 expression in macrophages (Cushing et al. (1990) Proc Natl Acad Sci 87:5134-5138; Yoshida et al. (1998) Arterioscler Thromb Vasc Biol 18:794-802; Steinberg (1997) J Biol Chem 272:20963-20966). SRA and the other scavenger receptors can bind Ox-LDL and enhance uptake of lipoprotein particles.

Mononuclear phagocytes enter the intima, differentiate into macrophages, and ingest modified lipids including Ox-LDL. In most cell types, cholesterol content is tightly controlled by feedback regulation of LDL receptors and biosynthetic enzymes (Brown and Goldstein (1986) Science 232:34-47). In macrophages, however, the additional scavenger receptors lead to unregulated uptake of cholesterol (Brown and Goldstein (1983) Annu Rev Biochem 52:223-261) and accumulation of multiple intracellular lipid droplets producing a "foam cell" phenotype. Cholesterol-engorged and dead macrophages contribute most of the mass of early "fatty streak" plaques and typical "advanced" lesions of diseased arteries. Numerous studies have described a variety of foam cell responses that contribute to growth and rupture of atherosclerotic vessel wall plaques. These responses include production of multiple growth factors and cytokines, which promote proliferation and adherence of neighboring cells; chemokines, which further attract circulating monocytes into the growing plaque; proteins, which cause remodeling of the extracellular matrix; and tissue factor, which can trigger thrombosis (Ross (1993) Nature 362:801-809; Quin et al. (1987) Proc Natl Acad Sci 84:2995-2998). Thus, cholesterol-loaded macrophages which occur in abundance in most stages of the atherosclerotic plaque formation contribute to inception of the atheroscerotic process and to eventual plaque rupture and occlusive thrombus.

During Ox-LDL uptake, macrophages produce cytokines and growth factors that elicit further cellular events that modulate atherogenesis such as smooth muscle cell proliferation and production of extracellular matrix. Additionally, these macrophages may activate genes involved in inflammation

including inducible nitric oxide synthase. Thus, genes differentially expressed during foam cell formation may reasonably be expected to be markers of the atherosclerotic process.

Association of NTRAN polynucleotides with Parkinson's Disease

Several genes have been identified as showing linkage to autosomal dominant forms of Parkinson's Disease (PD). PD is a common neurodegenerative disorder causing bradykinesia, resting tremor, muscular rigidity, and postural instability. Cytoplasmic eosinophilic inclusions called Lewy bodies, and neuronal loss especially in the substantia nigra pars compacta, are pathological hallmarks of PD (Valente, E.M. et al (2001) Am. J. Hum. Genet. 68:895-900). Lewy body Parkinson disease has been thought to be a specific autosomal dominant disorder (Wakabayashi, K. et al. (1998) Acta Neuropath. 96:207-210). Juvenile parkinsonism may be a specific autosomal recessive disorder (Matsumine, H. et al. (1997) Am. J. Hum. Genet. 60: 588-596, 1997). (Online Mendelian Inheritance in Man, OMIM. Johns Hopkins University, Baltimore, MD. MIM Number: 168600: Sept. 9, 2002: . World Wide Web URL: http://www.ncbi.nlm.nih.gov/omim/)

The discovery of new neurotransmission-associated proteins, and the polynucleotides encoding them, satisfies a need in the art by providing new compositions which are useful in the diagnosis, prevention, and treatment of autoimmune/inflammatory disorders, cardiovascular disorders, neurological diseases, developmental disorders, and cell proliferative diseases and disorders, including cancer, and in the assessment of the effects of exogenous compounds on the expression of nucleic acid and amino acid sequences of neurotransmission-associated proteins.

There is a need in the art for new compositions, including nucleic acids and proteins, for the diagnosis, prevention, and treatment of autoimmune/inflammatory, cardiovascular, neurological, developmental, cell proliferative, transport, psychiatric, metabolic, and endocrine disorders.

SUMMARY OF THE INVENTION

Various embodiments of the invention provide purified polypeptides, neurotransmission-associated proteins, referred to collectively as 'NTRAN' and individually as 'NTRAN-1,' 'NTRAN-2,' 'NTRAN-3,' 'NTRAN-4,' 'NTRAN-5,' 'NTRAN-6,' 'NTRAN-7,' 'NTRAN-8,' 'NTRAN-9,' 'NTRAN-10,' 'NTRAN-11,' 'NTRAN-12,' 'NTRAN-13,' 'NTRAN-14,' 'NTRAN-15,' 'NTRAN-16,' 'NTRAN-17,' 'NTRAN-18,' 'NTRAN-19,' 'NTRAN-20,' 'NTRAN-21,' and 'NTRAN-22' and methods for using these proteins and their encoding polynucleotides for the detection, diagnosis, and treatment of diseases and medical conditions. Embodiments also provide methods for utilizing the purified neurotransmission-associated proteins and/or their encoding

polynucleotides for facilitating the drug discovery process, including determination of efficacy, dosage, toxicity, and pharmacology. Related embodiments provide methods for utilizing the purified neurotransmission-associated proteins and/or their encoding polynucleotides for investigating the pathogenesis of diseases and medical conditions.

An embodiment provides an isolated polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-22, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical or at least about 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-22, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-22, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-22. Another embodiment provides an isolated polypeptide comprising an amino acid sequence of SEQ ID NO:1-22.

Still another embodiment provides an isolated polynucleotide encoding a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-22, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical or at least about 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-22, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-22, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-22. In another embodiment, the polynucleotide encodes a polypeptide selected from the group consisting of SEQ ID NO:1-22. In an alternative embodiment, the polynucleotide is selected from the group consisting of SEQ ID NO:23-44.

Still another embodiment provides a recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide encoding a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-22, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical or at least about 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-22, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-22, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-22. Another embodiment provides a cell transformed with the recombinant polynucleotide. Yet another embodiment provides a transgenic organism comprising the recombinant polynucleotide.

Another embodiment provides a method for producing a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-22, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical or at least about 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-22, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-22, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-22. The method comprises a) culturing a cell under conditions suitable for expression of the polypeptide, wherein said cell is transformed with a recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide encoding the polypeptide, and b) recovering the polypeptide so expressed.

Yet another embodiment provides an isolated antibody which specifically binds to a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-22, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical or at least about 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-22, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-22, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-22.

Still yet another embodiment provides an isolated polynucleotide selected from the group consisting of a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:23-44, b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical or at least about 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:23-44, c) a polynucleotide complementary to the polynucleotide of a), d) a polynucleotide complementary to the polynucleotide of b), and e) an RNA equivalent of a)-d). In other embodiments, the polynucleotide can comprise at least about 20, 30, 40, 60, 80, or 100 contiguous nucleotides.

Yet another embodiment provides a method for detecting a target polynucleotide in a sample, said target polynucleotide being selected from the group consisting of a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:23-44, b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical or at least about 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:23-44, c) a

polynucleotide complementary to the polynucleotide of a), d) a polynucleotide complementary to the polynucleotide of b), and e) an RNA equivalent of a)-d). The method comprises a) hybridizing the sample with a probe comprising at least 20 contiguous nucleotides comprising a sequence complementary to said target polynucleotide in the sample, and which probe specifically hybridizes to said target polynucleotide, under conditions whereby a hybridization complex is formed between said probe and said target polynucleotide or fragments thereof, and b) detecting the presence or absence of said hybridization complex. In a related embodiment, the method can include detecting the amount of the hybridization complex. In still other embodiments, the probe can comprise at least about 20, 30, 40, 60, 80, or 100 contiguous nucleotides.

Still yet another embodiment provides a method for detecting a target polynucleotide in a sample, said target polynucleotide being selected from the group consisting of a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:23-44, b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical or at least about 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:23-44, c) a polynucleotide complementary to the polynucleotide of a), d) a polynucleotide complementary to the polynucleotide of b), and e) an RNA equivalent of a)-d). The method comprises a) amplifying said target polynucleotide or fragment thereof using polymerase chain reaction amplification, and b) detecting the presence or absence of said amplified target polynucleotide or fragment thereof. In a related embodiment, the method can include detecting the amount of the amplified target polynucleotide or fragment thereof.

Another embodiment provides a composition comprising an effective amount of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-22, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical or at least about 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-22, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-22, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-22, and a pharmaceutically acceptable excipient. In one embodiment, the composition can comprise an amino acid sequence selected from the group consisting of SEQ ID NO:1-22. Other embodiments provide a method of treating a disease or condition associated with decreased or abnormal expression of functional NTRAN, comprising administering to a patient in need of such treatment the composition.

Yet another embodiment provides a method for screening a compound for effectiveness as an agonist of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-22, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical or at least about 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-22, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-22, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-22. The method comprises a) exposing a sample comprising the polypeptide to a compound, and b) detecting agonist activity in the sample. Another embodiment provides a composition comprising an agonist compound identified by the method and a pharmaceutically acceptable excipient. Yet another embodiment provides a method of treating a disease or condition associated with decreased expression of functional NTRAN, comprising administering to a patient in need of such treatment the composition.

Still yet another embodiment provides a method for screening a compound for effectiveness as an antagonist of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-22, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical or at least about 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-22, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-22, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-22. The method comprises a) exposing a sample comprising the polypeptide to a compound, and b) detecting antagonist activity in the sample. Another embodiment provides a composition comprising an antagonist compound identified by the method and a pharmaceutically acceptable excipient. Yet another embodiment provides a method of treating a disease or condition associated with overexpression of functional NTRAN, comprising administering to a patient in need of such treatment the composition.

Another embodiment provides a method of screening for a compound that specifically binds to a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-22, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical or at least about 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-22, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ

ID NO:1-22, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-22. The method comprises a) combining the polypeptide with at least one test compound under suitable conditions, and b) detecting binding of the polypeptide to the test compound, thereby identifying a compound that specifically binds to the polypeptide.

Yet another embodiment provides a method of screening for a compound that modulates the activity of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-22, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical or at least about 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-22, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-22, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-22. The method comprises a) combining the polypeptide with at least one test compound under conditions permissive for the activity of the polypeptide, b) assessing the activity of the polypeptide in the presence of the test compound, and c) comparing the activity of the polypeptide in the presence of the test compound with the activity of the polypeptide in the absence of the test compound, wherein a change in the activity of the polypeptide in the presence of the test compound is indicative of a compound that modulates the activity of the polypeptide.

Still yet another embodiment provides a method for screening a compound for effectiveness in altering expression of a target polynucleotide, wherein said target polynucleotide comprises a polynucleotide sequence selected from the group consisting of SEQ ID NO:23-44, the method comprising a) exposing a sample comprising the target polynucleotide to a compound, b) detecting altered expression of the target polynucleotide, and c) comparing the expression of the target polynucleotide in the presence of varying amounts of the compound and in the absence of the compound.

Another embodiment provides a method for assessing toxicity of a test compound, said method comprising a) treating a biological sample containing nucleic acids with the test compound; b) hybridizing the nucleic acids of the treated biological sample with a probe comprising at least 20 contiguous nucleotides of a polynucleotide selected from the group consisting of i) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:23-44, ii) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical or at

least about 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:23-44, iii) a polynucleotide having a sequence complementary to i), iv) a polynucleotide complementary to the polynucleotide of ii), and v) an RNA equivalent of i)-iv). Hybridization occurs under conditions whereby a specific hybridization complex is formed between said probe and a target polynucleotide in the biological sample, said target polynucleotide selected from the group consisting of i) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:23-44, ii) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical or at least about 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:23-44, iii) a polynucleotide complementary to the polynucleotide of i), iv) a polynucleotide complementary to the polynucleotide of ii), and v) an RNA equivalent of i)-iv). Alternatively, the target polynucleotide can comprise a fragment of a polynucleotide selected from the group consisting of i)-v) above; c) quantifying the amount of hybridization complex; and d) comparing the amount of hybridization complex in the treated biological sample with the amount of hybridization complex in an untreated biological sample, wherein a difference in the amount of hybridization complex in the treated biological sample is indicative of toxicity of the test compound.

BRIEF DESCRIPTION OF THE TABLES

Table 1 summarizes the nomenclature for full length polynucleotide and polypeptide embodiments of the invention.

Table 2 shows the GenBank identification number and annotation of the nearest GenBank homolog, and the PROTEOME database identification numbers and annotations of PROTEOME database homologs, for polypeptide embodiments of the invention. The probability scores for the matches between each polypeptide and its homolog(s) are also shown.

Table 3 shows structural features of polypeptide embodiments, including predicted motifs and domains, along with the methods, algorithms, and searchable databases used for analysis of the polypeptides.

Table 4 lists the cDNA and/or genomic DNA fragments which were used to assemble polynucleotide embodiments, along with selected fragments of the polynucleotides.

Table 5 shows representative cDNA libraries for polynucleotide embodiments.

Table 6 provides an appendix which describes the tissues and vectors used for construction of the cDNA libraries shown in Table 5.

Table 7 shows the tools, programs, and algorithms used to analyze polynucleotides and

polypeptides, along with applicable descriptions, references, and threshold parameters.

Table 8 shows single nucleotide polymorphisms found in polynucleotide sequences of the invention, along with allele frequencies in different human populations.

DESCRIPTION OF THE INVENTION

Before the present proteins, nucleic acids, and methods are described, it is understood that embodiments of the invention are not limited to the particular machines, instruments, materials, and methods described, as these may vary. It is also to be understood that the terminology used herein is for the purpose of describing particular embodiments only, and is not intended to limit the scope of the invention.

As used herein and in the appended claims, the singular forms "a," "an," and "the" include plural reference unless the context clearly dictates otherwise. Thus, for example, a reference to "a host cell" includes a plurality of such host cells, and a reference to "an antibody" is a reference to one or more antibodies and equivalents thereof known to those skilled in the art, and so forth.

Unless defined otherwise, all technical and scientific terms used herein have the same meanings as commonly understood by one of ordinary skill in the art to which this invention belongs. Although any machines, materials, and methods similar or equivalent to those described herein can be used to practice or test the present invention, the preferred machines, materials and methods are now described. All publications mentioned herein are cited for the purpose of describing and disclosing the cell lines, protocols, reagents and vectors which are reported in the publications and which might be used in connection with various embodiments of the invention. Nothing herein is to be construed as an admission that the invention is not entitled to antedate such disclosure by virtue of prior invention.

DEFINITIONS

"NTRAN" refers to the amino acid sequences of substantially purified NTRAN obtained from any species, particularly a mammalian species, including bovine, ovine, porcine, murine, equine, and human, and from any source, whether natural, synthetic, semi-synthetic, or recombinant.

The term "agonist" refers to a molecule which intensifies or mimics the biological activity of NTRAN. Agonists may include proteins, nucleic acids, carbohydrates, small molecules, or any other compound or composition which modulates the activity of NTRAN either by directly interacting with NTRAN or by acting on components of the biological pathway in which NTRAN participates.

An "allelic variant" is an alternative form of the gene encoding NTRAN. Allelic variants may result from at least one mutation in the nucleic acid sequence and may result in altered mRNAs or in

polypeptides whose structure or function may or may not be altered. A gene may have none, one, or many allelic variants of its naturally occurring form. Common mutational changes which give rise to allelic variants are generally ascribed to natural deletions, additions, or substitutions of nucleotides. Each of these types of changes may occur alone, or in combination with the others, one or more times in a given sequence.

"Altered" nucleic acid sequences encoding NTRAN include those sequences with deletions, insertions, or substitutions of different nucleotides, resulting in a polypeptide the same as NTRAN or a polypeptide with at least one functional characteristic of NTRAN. Included within this definition are polymorphisms which may or may not be readily detectable using a particular oligonucleotide probe of the polynucleotide encoding NTRAN, and improper or unexpected hybridization to allelic variants, with a locus other than the normal chromosomal locus for the polynucleotide encoding NTRAN. The encoded protein may also be "altered," and may contain deletions, insertions, or substitutions of amino acid residues which produce a silent change and result in a functionally equivalent NTRAN. Deliberate amino acid substitutions may be made on the basis of one or more similarities in polarity, charge, solubility, hydrophobicity, hydrophilicity, and/or the amphipathic nature of the residues, as long as the biological or immunological activity of NTRAN is retained. For example, negatively charged amino acids may include aspartic acid and glutamic acid, and positively charged amino acids may include lysine and arginine. Amino acids with uncharged polar side chains having similar hydrophilicity values may include: asparagine and glutamine; and serine and threonine. Amino acids with uncharged side chains having similar hydrophilicity values may include: leucine, isoleucine, and valine; glycine and alanine; and phenylalanine and tyrosine.

The terms "amino acid" and "amino acid sequence" can refer to an oligopeptide, a peptide, a polypeptide, or a protein sequence, or a fragment of any of these, and to naturally occurring or synthetic molecules. Where "amino acid sequence" is recited to refer to a sequence of a naturally occurring protein molecule, "amino acid sequence" and like terms are not meant to limit the amino acid sequence to the complete native amino acid sequence associated with the recited protein molecule.

"Amplification" relates to the production of additional copies of a nucleic acid. Amplification may be carried out using polymerase chain reaction (PCR) technologies or other nucleic acid amplification technologies well known in the art.

The term "antagonist" refers to a molecule which inhibits or attenuates the biological activity of NTRAN. Antagonists may include proteins such as antibodies, anticalins, nucleic acids, carbohydrates, small molecules, or any other compound or composition which modulates the activity of

NTRAN either by directly interacting with NTRAN or by acting on components of the biological pathway in which NTRAN participates.

The term "antibody" refers to intact immunoglobulin molecules as well as to fragments thereof, such as Fab, $F(ab')_2$, and Fv fragments, which are capable of binding an epitopic determinant. Antibodies that bind NTRAN polypeptides can be prepared using intact polypeptides or using fragments containing small peptides of interest as the immunizing antigen. The polypeptide or oligopeptide used to immunize an animal (e.g., a mouse, a rat, or a rabbit) can be derived from the translation of RNA, or synthesized chemically, and can be conjugated to a carrier protein if desired. Commonly used carriers that are chemically coupled to peptides include bovine serum albumin, thyroglobulin, and keyhole limpet hemocyanin (KLH). The coupled peptide is then used to immunize the animal.

The term "antigenic determinant" refers to that region of a molecule (i.e., an epitope) that makes contact with a particular antibody. When a protein or a fragment of a protein is used to immunize a host animal, numerous regions of the protein may induce the production of antibodies which bind specifically to antigenic determinants (particular regions or three-dimensional structures on the protein). An antigenic determinant may compete with the intact antigen (i.e., the immunogen used to elicit the immune response) for binding to an antibody.

The term "aptamer" refers to a nucleic acid or oligonucleotide molecule that binds to a specific molecular target. Aptamers are derived from an *in vitro* evolutionary process (e.g., SELEX (Systematic Evolution of Ligands by EXponential Enrichment), described in U.S. Patent No. 5,270,163), which selects for target-specific aptamer sequences from large combinatorial libraries. Aptamer compositions may be double-stranded or single-stranded, and may include deoxyribonucleotides, ribonucleotides, nucleotide derivatives, or other nucleotide-like molecules. The nucleotide components of an aptamer may have modified sugar groups (e.g., the 2'-OH group of a ribonucleotide may be replaced by 2'-F or $2'-NH_2$), which may improve a desired property, e.g., resistance to nucleases or longer lifetime in blood. Aptamers may be conjugated to other molecules, e.g., a high molecular weight carrier to slow clearance of the aptamer from the circulatory system. Aptamers may be specifically cross-linked to their cognate ligands, e.g., by photo-activation of a cross-linker (Brody, E.N. and L. Gold (2000) J. Biotechnol. 74:5-13).

The term "intramer" refers to an aptamer which is expressed *in vivo*. For example, a vaccinia virus-based RNA expression system has been used to express specific RNA aptamers at high levels in the cytoplasm of leukocytes (Blind, M. et al. (1999) Proc. Natl. Acad. Sci. USA

96:3606-3610).

The term "spiegelmer" refers to an aptamer which includes L-DNA, L-RNA, or other left-handed nucleotide derivatives or nucleotide-like molecules. Aptamers containing left-handed nucleotides are resistant to degradation by naturally occurring enzymes, which normally act on substrates containing right-handed nucleotides.

The term "antisense" refers to any composition capable of base-pairing with the "sense" (coding) strand of a polynucleotide having a specific nucleic acid sequence. Antisense compositions may include DNA; RNA; peptide nucleic acid (PNA); oligonucleotides having modified backbone linkages such as phosphorothioates, methylphosphonates, or benzylphosphonates; oligonucleotides having modified sugar groups such as 2'-methoxyethyl sugars or 2'-methoxyethoxy sugars; or oligonucleotides having modified bases such as 5-methyl cytosine, 2'-deoxyuracil, or 7-deaza-2'-deoxyguanosine. Antisense molecules may be produced by any method including chemical synthesis or transcription. Once introduced into a cell, the complementary antisense molecule base-pairs with a naturally occurring nucleic acid sequence produced by the cell to form duplexes which block either transcription or translation. The designation "negative" or "minus" can refer to the antisense strand, and the designation "positive" or "plus" can refer to the sense strand of a reference DNA molecule.

The term "biologically active" refers to a protein having structural, regulatory, or biochemical functions of a naturally occurring molecule. Likewise, "immunologically active" or "immunogenic" refers to the capability of the natural, recombinant, or synthetic NTRAN, or of any oligopeptide thereof, to induce a specific immune response in appropriate animals or cells and to bind with specific antibodies.

"Complementary" describes the relationship between two single-stranded nucleic acid sequences that anneal by base-pairing. For example, 5'-AGT-3' pairs with its complement, 3'-TCA-5'.

A "composition comprising a given polynucleotide" and a "composition comprising a given polypeptide" can refer to any composition containing the given polynucleotide or polypeptide. The composition may comprise a dry formulation or an aqueous solution. Compositions comprising polynucleotides encoding NTRAN or fragments of NTRAN may be employed as hybridization probes. The probes may be stored in freeze-dried form and may be associated with a stabilizing agent such as a carbohydrate. In hybridizations, the probe may be deployed in an aqueous solution containing salts (e.g., NaCl), detergents (e.g., sodium dodecyl sulfate; SDS), and other components (e.g., Denhardt's solution, dry milk, salmon sperm DNA, etc.).

"Consensus sequence" refers to a nucleic acid sequence which has been subjected to repeated DNA sequence analysis to resolve uncalled bases, extended using the XL-PCR kit (Applied Biosystems, Foster City CA) in the 5' and/or the 3' direction, and resequenced, or which has been assembled from one or more overlapping cDNA, EST, or genomic DNA fragments using a computer program for fragment assembly, such as the GELVIEW fragment assembly system (Accelrys, Burlington MA) or Phrap (University of Washington, Seattle WA). Some sequences have been both extended and assembled to produce the consensus sequence.

"Conservative amino acid substitutions" are those substitutions that are predicted to least interfere with the properties of the original protein, i.e., the structure and especially the function of the protein is conserved and not significantly changed by such substitutions. The table below shows amino acids which may be substituted for an original amino acid in a protein and which are regarded as conservative amino acid substitutions.

Original Residue	Conservative Substitution
Ala	Gly, Ser
Arg	His, Lys
Asn	Asp, Gln, His
Asp	Asn, Glu
Cys	Ala, Ser
Gln	Asn, Glu, His
Glu	Asp, Gln, His
Gly	Ala
His	Asn, Arg, Gln, Glu
Ile	Leu, Val
Leu	Ile, Val
Lys	Arg, Gln, Glu
Met	Leu, Ile
Phe	His, Met, Leu, Trp, Tyr
Ser	Cys, Thr
Thr	Ser, Val
Trp	Phe, Tyr
Tyr	His, Phe, Trp
Val	Ile, Leu, Thr

Conservative amino acid substitutions generally maintain (a) the structure of the polypeptide backbone in the area of the substitution, for example, as a beta sheet or alpha helical conformation, (b) the charge or hydrophobicity of the molecule at the site of the substitution, and/or (c) the bulk of the side chain.

A "deletion" refers to a change in the amino acid or nucleotide sequence that results in the absence of one or more amino acid residues or nucleotides.

The term "derivative" refers to a chemically modified polynucleotide or polypeptide. Chemical modifications of a polynucleotide can include, for example, replacement of hydrogen by an alkyl, acyl, hydroxyl, or amino group. A derivative polynucleotide encodes a polypeptide which retains at least one biological or immunological function of the natural molecule. A derivative polypeptide is one modified by glycosylation, pegylation, or any similar process that retains at least one biological or immunological function of the polypeptide from which it was derived.

A "detectable label" refers to a reporter molecule or enzyme that is capable of generating a measurable signal and is covalently or noncovalently joined to a polynucleotide or polypeptide.

"Differential expression" refers to increased or upregulated; or decreased, downregulated, or absent gene or protein expression, determined by comparing at least two different samples. Such comparisons may be carried out between, for example, a treated and an untreated sample, or a diseased and a normal sample.

"Exon shuffling" refers to the recombination of different coding regions (exons). Since an exon may represent a structural or functional domain of the encoded protein, new proteins may be assembled through the novel reassortment of stable substructures, thus allowing acceleration of the evolution of new protein functions.

A "fragment" is a unique portion of NTRAN or a polynucleotide encoding NTRAN which can be identical in sequence to, but shorter in length than, the parent sequence. A fragment may comprise up to the entire length of the defined sequence, minus one nucleotide/amino acid residue. For example, a fragment may comprise from about 5 to about 1000 contiguous nucleotides or amino acid residues. A fragment used as a probe, primer, antigen, therapeutic molecule, or for other purposes, may be at least 5, 10, 15, 16, 20, 25, 30, 40, 50, 60, 75, 100, 150, 250 or at least 500 contiguous nucleotides or amino acid residues in length. Fragments may be preferentially selected from certain regions of a molecule. For example, a polypeptide fragment may comprise a certain length of contiguous amino acids selected from the first 250 or 500 amino acids (or first 25% or 50%) of a polypeptide as shown in a certain defined sequence. Clearly these lengths are exemplary, and any length that is supported by the specification, including the Sequence Listing, tables, and figures, may be encompassed by the present embodiments.

A fragment of SEQ ID NO:23-44 can comprise a region of unique polynucleotide sequence that specifically identifies SEQ ID NO:23-44, for example, as distinct from any other sequence in the genome from which the fragment was obtained. A fragment of SEQ ID NO:23-44 can be employed in one or more embodiments of methods of the invention, for example, in hybridization and

amplification technologies and in analogous methods that distinguish SEQ ID NO:23-44 from related polynucleotides. The precise length of a fragment of SEQ ID NO:23-44 and the region of SEQ ID NO:23-44 to which the fragment corresponds are routinely determinable by one of ordinary skill in the art based on the intended purpose for the fragment.

A fragment of SEQ ID NO:1-22 is encoded by a fragment of SEQ ID NO:23-44. A fragment of SEQ ID NO:1-22 can comprise a region of unique amino acid sequence that specifically identifies SEQ ID NO:1-22. For example, a fragment of SEQ ID NO:1-22 can be used as an immunogenic peptide for the development of antibodies that specifically recognize SEQ ID NO:1-22. The precise length of a fragment of SEQ ID NO:1-22 and the region of SEQ ID NO:1-22 to which the fragment corresponds can be determined based on the intended purpose for the fragment using one or more analytical methods described herein or otherwise known in the art.

A "full length" polynucleotide is one containing at least a translation initiation codon (e.g., methionine) followed by an open reading frame and a translation termination codon. A "full length" polynucleotide sequence encodes a "full length" polypeptide sequence.

"Homology" refers to sequence similarity or, alternatively, sequence identity, between two or more polynucleotide sequences or two or more polypeptide sequences.

The terms "percent identity" and "% identity," as applied to polynucleotide sequences, refer to the percentage of identical residue matches between at least two polynucleotide sequences aligned using a standardized algorithm. Such an algorithm may insert, in a standardized and reproducible way, gaps in the sequences being compared in order to optimize alignment between two sequences, and therefore achieve a more meaningful comparison of the two sequences.

Percent identity between polynucleotide sequences may be determined using one or more computer algorithms or programs known in the art or described herein. For example, percent identity can be determined using the default parameters of the CLUSTAL V algorithm as incorporated into the MEGALIGN version 3.12e sequence alignment program. This program is part of the LASERGENE software package, a suite of molecular biological analysis programs (DNASTAR, Madison WI). CLUSTAL V is described in Higgins, D.G. and P.M. Sharp (1989; CABIOS 5:151-153) and in Higgins, D.G. et al. (1992; CABIOS 8:189-191). For pairwise alignments of polynucleotide sequences, the default parameters are set as follows: Ktuple=2, gap penalty=5, window=4, and "diagonals saved"=4. The "weighted" residue weight table is selected as the default.

Alternatively, a suite of commonly used and freely available sequence comparison algorithms which can be used is provided by the National Center for Biotechnology Information (NCBI) Basic

Local Alignment Search Tool (BLAST) (Altschul, S.F. et al. (1990) J. Mol. Biol. 215:403-410), which is available from several sources, including the NCBI, Bethesda, MD, and on the Internet at http://www.ncbi.nlm.nih.gov/BLAST/. The BLAST software suite includes various sequence analysis programs including "blastn," that is used to align a known polynucleotide sequence with other polynucleotide sequences from a variety of databases. Also available is a tool called "BLAST 2 Sequences" that is used for direct pairwise comparison of two nucleotide sequences. "BLAST 2 Sequences" can be accessed and used interactively at http://www.ncbi.nlm.nih.gov/gorf/bl2.html. The "BLAST 2 Sequences" tool can be used for both blastn and blastp (discussed below). BLAST programs are commonly used with gap and other parameters set to default settings. For example, to compare two nucleotide sequences, one may use blastn with the "BLAST 2 Sequences" tool Version 2.0.12 (April-21-2000) set at default parameters. Such default parameters may be, for example:

Matrix: BLOSUM62

Reward for match: 1

Penalty for mismatch: -2

Open Gap: 5 and Extension Gap: 2 penalties

Gap x drop-off: 50

Expect: 10

Word Size: 11

Filter: on

Percent identity may be measured over the length of an entire defined sequence, for example, as defined by a particular SEQ ID number, or may be measured over a shorter length, for example, over the length of a fragment taken from a larger, defined sequence, for instance, a fragment of at least 20, at least 30, at least 40, at least 50, at least 70, at least 100, or at least 200 contiguous nucleotides. Such lengths are exemplary only, and it is understood that any fragment length supported by the sequences shown herein, in the tables, figures, or Sequence Listing, may be used to describe a length over which percentage identity may be measured.

Nucleic acid sequences that do not show a high degree of identity may nevertheless encode similar amino acid sequences due to the degeneracy of the genetic code. It is understood that changes in a nucleic acid sequence can be made using this degeneracy to produce multiple nucleic acid sequences that all encode substantially the same protein.

The phrases "percent identity" and "% identity," as applied to polypeptide sequences, refer to the percentage of identical residue matches between at least two polypeptide sequences aligned using

a standardized algorithm. Methods of polypeptide sequence alignment are well-known. Some alignment methods take into account conservative amino acid substitutions. Such conservative substitutions, explained in more detail above, generally preserve the charge and hydrophobicity at the site of substitution, thus preserving the structure (and therefore function) of the polypeptide. The phrases "percent similarity" and "% similarity," as applied to polypeptide sequences, refer to the percentage of residue matches, including identical residue matches and conservative substitutions, between at least two polypeptide sequences aligned using a standardized algorithm. In contrast, conservative substitutions are not included in the calculation of percent identity between polypeptide sequences.

Percent identity between polypeptide sequences may be determined using the default parameters of the CLUSTAL V algorithm as incorporated into the MEGALIGN version 3.12e sequence alignment program (described and referenced above). For pairwise alignments of polypeptide sequences using CLUSTAL V, the default parameters are set as follows: Ktuple=1, gap penalty=3, window=5, and "diagonals saved"=5. The PAM250 matrix is selected as the default residue weight table.

Alternatively the NCBI BLAST software suite may be used. For example, for a pairwise comparison of two polypeptide sequences, one may use the "BLAST 2 Sequences" tool Version 2.0.12 (April-21-2000) with blastp set at default parameters. Such default parameters may be, for example:

Matrix: BLOSUM62
Open Gap: 11 and Extension Gap: 1 penalties
Gap x drop-off: 50
Expect: 10
Word Size: 3
Filter: on

Percent identity may be measured over the length of an entire defined polypeptide sequence, for example, as defined by a particular SEQ ID number, or may be measured over a shorter length, for example, over the length of a fragment taken from a larger, defined polypeptide sequence, for instance, a fragment of at least 15, at least 20, at least 30, at least 40, at least 50, at least 70 or at least 150 contiguous residues. Such lengths are exemplary only, and it is understood that any fragment length supported by the sequences shown herein, in the tables, figures or Sequence Listing, may be used to describe a length over which percentage identity may be measured.

"Human artificial chromosomes" (HACs) are linear microchromosomes which may contain DNA sequences of about 6 kb to 10 Mb in size and which contain all of the elements required for chromosome replication, segregation and maintenance.

The term "humanized antibody" refers to an antibody molecule in which the amino acid sequence in the non-antigen binding regions has been altered so that the antibody more closely resembles a human antibody, and still retains its original binding ability.

"Hybridization" refers to the process by which a polynucleotide strand anneals with a complementary strand through base pairing under defined hybridization conditions. Specific hybridization is an indication that two nucleic acid sequences share a high degree of complementarity. Specific hybridization complexes form under permissive annealing conditions and remain hybridized after the "washing" step(s). The washing step(s) is particularly important in determining the stringency of the hybridization process, with more stringent conditions allowing less non-specific binding, i.e., binding between pairs of nucleic acid strands that are not perfectly matched. Permissive conditions for annealing of nucleic acid sequences are routinely determinable by one of ordinary skill in the art and may be consistent among hybridization experiments, whereas wash conditions may be varied among experiments to achieve the desired stringency, and therefore hybridization specificity. Permissive annealing conditions occur, for example, at 68°C in the presence of about 6 x SSC, about 1% (w/v) SDS, and about 100 μg/ml sheared, denatured salmon sperm DNA.

Generally, stringency of hybridization is expressed, in part, with reference to the temperature under which the wash step is carried out. Such wash temperatures are typically selected to be about 5°C to 20°C lower than the thermal melting point (T_m) for the specific sequence at a defined ionic strength and pH. The T_m is the temperature (under defined ionic strength and pH) at which 50% of the target sequence hybridizes to a perfectly matched probe. An equation for calculating T_m and conditions for nucleic acid hybridization are well known and can be found in Sambrook, J. and D.W. Russell (2001; Molecular Cloning: A Laboratory Manual, 3rd ed., vol. 1-3, Cold Spring Harbor Press, Cold Spring Harbor NY, ch. 9).

High stringency conditions for hybridization between polynucleotides of the present invention include wash conditions of 68°C in the presence of about 0.2 x SSC and about 0.1% SDS, for 1 hour. Alternatively, temperatures of about 65°C, 60°C, 55°C, or 42°C may be used. SSC concentration may be varied from about 0.1 to 2 x SSC, with SDS being present at about 0.1%. Typically, blocking reagents are used to block non-specific hybridization. Such blocking reagents include, for instance, sheared and denatured salmon sperm DNA at about 100-200 μg/ml. Organic solvent, such as

formamide at a concentration of about 35-50% v/v, may also be used under particular circumstances, such as for RNA:DNA hybridizations. Useful variations on these wash conditions will be readily apparent to those of ordinary skill in the art. Hybridization, particularly under high stringency conditions, may be suggestive of evolutionary similarity between the nucleotides. Such similarity is strongly indicative of a similar role for the nucleotides and their encoded polypeptides.

The term "hybridization complex" refers to a complex formed between two nucleic acids by virtue of the formation of hydrogen bonds between complementary bases. A hybridization complex may be formed in solution (e.g., C_0t or R_0t analysis) or formed between one nucleic acid present in solution and another nucleic acid immobilized on a solid support (e.g., paper, membranes, filters, chips, pins or glass slides, or any other appropriate substrate to which cells or their nucleic acids have been fixed).

The words "insertion" and "addition" refer to changes in an amino acid or polynucleotide sequence resulting in the addition of one or more amino acid residues or nucleotides, respectively.

"Immune response" can refer to conditions associated with inflammation, trauma, immune disorders, or infectious or genetic disease, etc. These conditions can be characterized by expression of various factors, e.g., cytokines, chemokines, and other signaling molecules, which may affect cellular and systemic defense systems.

An "immunogenic fragment" is a polypeptide or oligopeptide fragment of NTRAN which is capable of eliciting an immune response when introduced into a living organism, for example, a mammal. The term "immunogenic fragment" also includes any polypeptide or oligopeptide fragment of NTRAN which is useful in any of the antibody production methods disclosed herein or known in the art.

The term "microarray" refers to an arrangement of a plurality of polynucleotides, polypeptides, antibodies, or other chemical compounds on a substrate.

The terms "element" and "array element" refer to a polynucleotide, polypeptide, antibody, or other chemical compound having a unique and defined position on a microarray.

The term "modulate" refers to a change in the activity of NTRAN. For example, modulation may cause an increase or a decrease in protein activity, binding characteristics, or any other biological, functional, or immunological properties of NTRAN.

The phrases "nucleic acid" and "nucleic acid sequence" refer to a nucleotide, oligonucleotide, polynucleotide, or any fragment thereof. These phrases also refer to DNA or RNA of genomic or synthetic origin which may be single-stranded or double-stranded and may represent the sense or the

antisense strand, to peptide nucleic acid (PNA), or to any DNA-like or RNA-like material.

"Operably linked" refers to the situation in which a first nucleic acid sequence is placed in a functional relationship with a second nucleic acid sequence. For instance, a promoter is operably linked to a coding sequence if the promoter affects the transcription or expression of the coding sequence. Operably linked DNA sequences may be in close proximity or contiguous and, where necessary to join two protein coding regions, in the same reading frame.

"Peptide nucleic acid" (PNA) refers to an antisense molecule or anti-gene agent which comprises an oligonucleotide of at least about 5 nucleotides in length linked to a peptide backbone of amino acid residues ending in lysine. The terminal lysine confers solubility to the composition. PNAs preferentially bind complementary single stranded DNA or RNA and stop transcript elongation, and may be pegylated to extend their lifespan in the cell.

"Post-translational modification" of an NTRAN may involve lipidation, glycosylation, phosphorylation, acetylation, racemization, proteolytic cleavage, and other modifications known in the art. These processes may occur synthetically or biochemically. Biochemical modifications will vary by cell type depending on the enzymatic milieu of NTRAN.

"Probe" refers to nucleic acids encoding NTRAN, their complements, or fragments thereof, which are used to detect identical, allelic or related nucleic acids. Probes are isolated oligonucleotides or polynucleotides attached to a detectable label or reporter molecule. Typical labels include radioactive isotopes, ligands, chemiluminescent agents, and enzymes. "Primers" are short nucleic acids, usually DNA oligonucleotides, which may be annealed to a target polynucleotide by complementary base-pairing. The primer may then be extended along the target DNA strand by a DNA polymerase enzyme. Primer pairs can be used for amplification (and identification) of a nucleic acid, e.g., by the polymerase chain reaction (PCR).

Probes and primers as used in the present invention typically comprise at least 15 contiguous nucleotides of a known sequence. In order to enhance specificity, longer probes and primers may also be employed, such as probes and primers that comprise at least 20, 25, 30, 40, 50, 60, 70, 80, 90, 100, or at least 150 consecutive nucleotides of the disclosed nucleic acid sequences. Probes and primers may be considerably longer than these examples, and it is understood that any length supported by the specification, including the tables, figures, and Sequence Listing, may be used.

Methods for preparing and using probes and primers are described in, for example, Sambrook, J. and D.W. Russell (2001; Molecular Cloning: A Laboratory Manual, 3rd ed., vol. 1-3, Cold Spring Harbor Press, Cold Spring Harbor NY), Ausubel, F.M. et al. (1999; Short Protocols in Molecular

Biology, 4th ed., John Wiley & Sons, New York NY), and Innis, M. et al. (1990; PCR Protocols, A Guide to Methods and Applications, Academic Press, San Diego CA). PCR primer pairs can be derived from a known sequence, for example, by using computer programs intended for that purpose such as Primer (Version 0.5, 1991, Whitehead Institute for Biomedical Research, Cambridge MA).

Oligonucleotides for use as primers are selected using software known in the art for such purpose. For example, OLIGO 4.06 software is useful for the selection of PCR primer pairs of up to 100 nucleotides each, and for the analysis of oligonucleotides and larger polynucleotides of up to 5,000 nucleotides from an input polynucleotide sequence of up to 32 kilobases. Similar primer selection programs have incorporated additional features for expanded capabilities. For example, the PrimOU primer selection program (available to the public from the Genome Center at University of Texas South West Medical Center, Dallas TX) is capable of choosing specific primers from megabase sequences and is thus useful for designing primers on a genome-wide scope. The Primer3 primer selection program (available to the public from the Whitehead Institute/MIT Center for Genome Research, Cambridge MA) allows the user to input a "mispriming library," in which sequences to avoid as primer binding sites are user-specified. Primer3 is useful, in particular, for the selection of oligonucleotides for microarrays. (The source code for the latter two primer selection programs may also be obtained from their respective sources and modified to meet the user's specific needs.) The PrimeGen program (available to the public from the UK Human Genome Mapping Project Resource Centre, Cambridge UK) designs primers based on multiple sequence alignments, thereby allowing selection of primers that hybridize to either the most conserved or least conserved regions of aligned nucleic acid sequences. Hence, this program is useful for identification of both unique and conserved oligonucleotides and polynucleotide fragments. The oligonucleotides and polynucleotide fragments identified by any of the above selection methods are useful in hybridization technologies, for example, as PCR or sequencing primers, microarray elements, or specific probes to identify fully or partially complementary polynucleotides in a sample of nucleic acids. Methods of oligonucleotide selection are not limited to those described above.

A "recombinant nucleic acid" is a nucleic acid that is not naturally occurring or has a sequence that is made by an artificial combination of two or more otherwise separated segments of sequence. This artificial combination is often accomplished by chemical synthesis or, more commonly, by the artificial manipulation of isolated segments of nucleic acids, e.g., by genetic engineering techniques such as those described in Sambrook and Russell (*supra*). The term recombinant includes nucleic acids that have been altered solely by addition, substitution, or deletion of a portion of the

nucleic acid. Frequently, a recombinant nucleic acid may include a nucleic acid sequence operably linked to a promoter sequence. Such a recombinant nucleic acid may be part of a vector that is used, for example, to transform a cell.

Alternatively, such recombinant nucleic acids may be part of a viral vector, e.g., based on a vaccinia virus, that could be use to vaccinate a mammal wherein the recombinant nucleic acid is expressed, inducing a protective immunological response in the mammal.

A "regulatory element" refers to a nucleic acid sequence usually derived from untranslated regions of a gene and includes enhancers, promoters, introns, and 5' and 3' untranslated regions (UTRs). Regulatory elements interact with host or viral proteins which control transcription, translation, or RNA stability.

"Reporter molecules" are chemical or biochemical moieties used for labeling a nucleic acid, amino acid, or antibody. Reporter molecules include radionuclides; enzymes; fluorescent, chemiluminescent, or chromogenic agents; substrates; cofactors; inhibitors; magnetic particles; and other moieties known in the art.

An "RNA equivalent," in reference to a DNA molecule, is composed of the same linear sequence of nucleotides as the reference DNA molecule with the exception that all occurrences of the nitrogenous base thymine are replaced with uracil, and the sugar backbone is composed of ribose instead of deoxyribose.

The term "sample" is used in its broadest sense. A sample suspected of containing NTRAN, nucleic acids encoding NTRAN, or fragments thereof may comprise a bodily fluid; an extract from a cell, chromosome, organelle, or membrane isolated from a cell; a cell; genomic DNA, RNA, or cDNA, in solution or bound to a substrate; a tissue; a tissue print; etc.

The terms "specific binding" and "specifically binding" refer to that interaction between a protein or peptide and an agonist, an antibody, an antagonist, a small molecule, or any natural or synthetic binding composition. The interaction is dependent upon the presence of a particular structure of the protein, e.g., the antigenic determinant or epitope, recognized by the binding molecule. For example, if an antibody is specific for epitope "A," the presence of a polypeptide comprising the epitope A, or the presence of free unlabeled A, in a reaction containing free labeled A and the antibody will reduce the amount of labeled A that binds to the antibody.

The term "substantially purified" refers to nucleic acid or amino acid sequences that are removed from their natural environment and are isolated or separated, and are at least about 60% free, preferably at least about 75% free, and most preferably at least about 90% free from other

components with which they are naturally associated.

A "substitution" refers to the replacement of one or more amino acid residues or nucleotides by different amino acid residues or nucleotides, respectively.

"Substrate" refers to any suitable rigid or semi-rigid support including membranes, filters, chips, slides, wafers, fibers, magnetic or nonmagnetic beads, gels, tubing, plates, polymers, microparticles and capillaries. The substrate can have a variety of surface forms, such as wells, trenches, pins, channels and pores, to which polynucleotides or polypeptides are bound.

A "transcript image" or "expression profile" refers to the collective pattern of gene expression by a particular cell type or tissue under given conditions at a given time.

"Transformation" describes a process by which exogenous DNA is introduced into a recipient cell. Transformation may occur under natural or artificial conditions according to various methods well known in the art, and may rely on any known method for the insertion of foreign nucleic acid sequences into a prokaryotic or eukaryotic host cell. The method for transformation is selected based on the type of host cell being transformed and may include, but is not limited to, bacteriophage or viral infection, electroporation, heat shock, lipofection, and particle bombardment. The term "transformed cells" includes stably transformed cells in which the inserted DNA is capable of replication either as an autonomously replicating plasmid or as part of the host chromosome, as well as transiently transformed cells which express the inserted DNA or RNA for limited periods of time.

A "transgenic organism," as used herein, is any organism, including but not limited to animals and plants, in which one or more of the cells of the organism contains heterologous nucleic acid introduced by way of human intervention, such as by transgenic techniques well known in the art. The nucleic acid is introduced into the cell, directly or indirectly by introduction into a precursor of the cell, by way of deliberate genetic manipulation, such as by microinjection or by infection with a recombinant virus. In another embodiment, the nucleic acid can be introduced by infection with a recombinant viral vector, such as a lentiviral vector (Lois, C. et al. (2002) Science 295:868-872). The term genetic manipulation does not include classical cross-breeding, or *in vitro* fertilization, but rather is directed to the introduction of a recombinant DNA molecule. The transgenic organisms contemplated in accordance with the present invention include bacteria, cyanobacteria, fungi, plants and animals. The isolated DNA of the present invention can be introduced into the host by methods known in the art, for example infection, transfection, transformation or transconjugation. Techniques for transferring the DNA of the present invention into such organisms are widely known and provided in references such as Sambrook and Russell (*supra*).

A "variant" of a particular nucleic acid sequence is defined as a nucleic acid sequence having at least 40% sequence identity to the particular nucleic acid sequence over a certain length of one of the nucleic acid sequences using blastn with the "BLAST 2 Sequences" tool Version 2.0.9 (May-07-1999) set at default parameters. Such a pair of nucleic acids may show, for example, at least 50%, at least 60%, at least 70%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% or greater sequence identity over a certain defined length. A variant may be described as, for example, an "allelic" (as defined above), "splice," "species," or "polymorphic" variant. A splice variant may have significant identity to a reference molecule, but will generally have a greater or lesser number of polynucleotides due to alternate splicing of exons during mRNA processing. The corresponding polypeptide may possess additional functional domains or lack domains that are present in the reference molecule. Species variants are polynucleotides that vary from one species to another. The resulting polypeptides will generally have significant amino acid identity relative to each other. A polymorphic variant is a variation in the polynucleotide sequence of a particular gene between individuals of a given species. Polymorphic variants also may encompass "single nucleotide polymorphisms" (SNPs) in which the polynucleotide sequence varies by one nucleotide base. The presence of SNPs may be indicative of, for example, a certain population, a disease state, or a propensity for a disease state.

A "variant" of a particular polypeptide sequence is defined as a polypeptide sequence having at least 40% sequence identity or sequence similarity to the particular polypeptide sequence over a certain length of one of the polypeptide sequences using blastp with the "BLAST 2 Sequences" tool Version 2.0.9 (May-07-1999) set at default parameters. Such a pair of polypeptides may show, for example, at least 50%, at least 60%, at least 70%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% or greater sequence identity or sequence similarity over a certain defined length of one of the polypeptides.

THE INVENTION

Various embodiments of the invention include new human neurotransmission-associated proteins (NTRAN), the polynucleotides encoding NTRAN, and the use of these compositions for the diagnosis, treatment, or prevention of autoimmune/inflammatory, cardiovascular, neurological, developmental, cell proliferative, transport, psychiatric, metabolic, and endocrine disorders.

Table 1 summarizes the nomenclature for the full length polynucleotide and polypeptide embodiments of the invention. Each polynucleotide and its corresponding polypeptide are correlated to a single Incyte project identification number (Incyte Project ID). Each polypeptide sequence is denoted by both a polypeptide sequence identification number (Polypeptide SEQ ID NO:) and an Incyte polypeptide sequence number (Incyte Polypeptide ID) as shown. Each polynucleotide sequence is denoted by both a polynucleotide sequence identification number (Polynucleotide SEQ ID NO:) and an Incyte polynucleotide consensus sequence number (Incyte Polynucleotide ID) as shown. Column 6 shows the Incyte ID numbers of physical, full length clones corresponding to the polypeptide and polynucleotide sequences of the invention. The full length clones encode polypeptides which have at least 95% sequence identity to the polypeptide sequences shown in column 3.

Table 2 shows sequences with homology to polypeptide embodiments of the invention as identified by BLAST analysis against the GenBank protein (genpept) database and the PROTEOME database. Columns 1 and 2 show the polypeptide sequence identification number (Polypeptide SEQ ID NO:) and the corresponding Incyte polypeptide sequence number (Incyte Polypeptide ID) for polypeptides of the invention. Column 3 shows the GenBank identification number (GenBank ID NO:) of the nearest GenBank homolog and the PROTEOME database identification numbers (PROTEOME ID NO:) of the nearest PROTEOME database homologs. Column 4 shows the probability scores for the matches between each polypeptide and its homolog(s). Column 5 shows the annotation of the GenBank and PROTEOME database homolog(s) along with relevant citations where applicable, all of which are expressly incorporated by reference herein.

Table 3 shows various structural features of the polypeptides of the invention. Columns 1 and 2 show the polypeptide sequence identification number (SEQ ID NO:) and the corresponding Incyte polypeptide sequence number (Incyte Polypeptide ID) for each polypeptide of the invention. Column 3 shows the number of amino acid residues in each polypeptide. Column 4 shows amino acid residues comprising signature sequences, domains, motifs, potential phosphorylation sites, and potential glycosylation sites. Column 5 shows analytical methods for protein structure/function analysis and in some cases, searchable databases to which the analytical methods were applied.

Together, Tables 2 and 3 summarize the properties of polypeptides of the invention, and these properties establish that the claimed polypeptides are neurotransmission-associated proteins. For example, For example, SEQ ID NO:5 is 68% identical, from residue E43 to residue D287, and is 60% identical, from residue M1 to G118 to murine gliacolin (GenBank ID g14278927) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability scores are

3.0e-86 and 5.6e-30, which indicate the probabilities of obtaining the observed polypeptide sequence alignments by chance. As determined by BLAST analysis using the PROTEOME database, SEQ ID NO:5 also has homology to murine and human C1q-related factors which have a collagenous region and a globular domain and have similarity to the C1q signature domain. Both murine and human C1q-related factors are highly expressed in brain areas that are involved in motor function (PROTEOME ID 429678|C1qrf and 567880|CRF). SEQ ID NO:5 also contains a C1q domain and a collagen triple helix repeat (20 copies) domain as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. (See Table 3.) Data from BLIMPS, MOTIFS, PROFILESCAN, and additional BLAST analyses provide further corroborative evidence that SEQ ID NO:5 is a C1q-related factor.

In an alternative example, SEQ ID NO:9 is 86% identical, from residue M1 to residue F1308, to Mus musculus cell recognition molecule CASPR4 (GenBank ID g12330704) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 0.0, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:9 also has homology to Neurexin 4 (contactin associated protein 1), a neuronal paranodal transmembrane receptor containing epidermal growth factor-like and neurexin motifs, as determined by BLAST analysis using the PROTEOME database. SEQ ID NO:9 also contains an EGF-like domain and a F5/8 type C domain as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. (See Table 3.) Data from BLIMPS and MOTIFS analyses provide further corroborative evidence that SEQ ID NO:9 is a cell recognition molecule.

In an alternative example, SEQ ID NO:13 is 100% identical, from residue M1 to residue S573, to human myelin-associated glycoprotein precursor (GenBank ID g307156) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 0.0, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:13 also has homology to proteins that are localized to the nervous system, are members of the sialoadhesin subgroup of immunoglobulin superfamily lectins, and are cell adhesion molecules, as determined by BLAST analysis using the PROTEOME database. SEQ ID NO:13 also contains an I-type IG domain, an Immunoglobulin domain, an Ig superfamily from SCOP domain and an Immunoglobulin C-2 type domain as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM, INCY and SMART databases of conserved protein families/domains. (See Table 3.) Data from BLIMPS, MOTIFS, and additional BLAST analyses

provide further corroborative evidence that SEQ ID NO:13 is a myelin cell adhesion molecule.

In an alternative example, SEQ ID NO:16 is 93% identical, from residue A32 to residue S412, to human ubiquilin (GenBank ID g18254511) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 4.6e-185, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:16 also has homology to proteins that are localized to the endoplasmic reticulum, nucleus, and cytoplasm, is associated with the 26S proteasome, binds to and promotes the accumulation of presenilin 1 and presenilin 2, and are localized to neurofibrillary tangles and Lewy bodies in brains affected by Alzheimer's disease and Parkinson's disease, and are annotated as ubiquilin 1 proteins, as determined by BLAST analysis using the PROTEOME database. SEQ ID NO:16 also contains UBA/TS-N and ubiquitin associated domains as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM and SMART database of conserved protein families/domains. (See Table 3.) Data from BLAST analyses against the PRODOM database, provide further corroborative evidence that SEQ ID NO:16 is a ubiquilin protein.

In an alternative example, SEQ ID NO:19 is 89% identical, from residue M1 to residue L621, to *Mus musculus* punc (g3068592) as determined by the Basic Local Alignment Search Tool (BLAST). The BLAST probability score is 0.0, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:19 also has homology to proteins that are localized to the plasma membrane, are members of the immunoglobulin superfamily, and are neuronal cell adhesion proteins, as determined by BLAST analysis using the PROTEOME database. SEQ ID NO:19 also contains fibronectin type 3 and immunoglobulin domains as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM, INCY, and SMART databases of conserved protein families/domains. (See Table 3.) Data from BLAST analyses against the PRODOM and DOMO databases, provide further corroborative evidence that SEQ ID NO:19 is a neuronal cell adhesion protein.

SEQ ID NO:1-4, SEQ ID NO:6-8, SEQ ID NO:10-12, SEQ ID NO:14-15, SEQ ID NO:17-18, and SEQ ID NO:20-22 were analyzed and annotated in a similar manner. The algorithms and parameters for the analysis of SEQ ID NO:1-22 are described in Table 7.

As shown in Table 4, the full length polynucleotide embodiments were assembled using cDNA sequences or coding (exon) sequences derived from genomic DNA, or any combination of these two types of sequences. Column 1 lists the polynucleotide sequence identification number (Polynucleotide SEQ ID NO:), the corresponding Incyte polynucleotide consensus sequence number (Incyte ID) for

each polynucleotide of the invention, and the length of each polynucleotide sequence in basepairs. Column 2 shows the nucleotide start (5′) and stop (3′) positions of the cDNA and/or genomic sequences used to assemble the full length polynucleotide embodiments, and of fragments of the polynucleotides which are useful, for example, in hybridization or amplification technologies that identify SEQ ID NO:23-44 or that distinguish between SEQ ID NO:23-44 and related polynucleotides.

The polynucleotide fragments described in Column 2 of Table 4 may refer specifically, for example, to Incyte cDNAs derived from tissue-specific cDNA libraries or from pooled cDNA libraries. Alternatively, the polynucleotide fragments described in column 2 may refer to GenBank cDNAs or ESTs which contributed to the assembly of the full length polynucleotides. In addition, the polynucleotide fragments described in column 2 may identify sequences derived from the ENSEMBL (The Sanger Centre, Cambridge, UK) database (*i.e.*, those sequences including the designation "ENST"). Alternatively, the polynucleotide fragments described in column 2 may be derived from the NCBI RefSeq Nucleotide Sequence Records Database (*i.e.*, those sequences including the designation "NM" or "NT") or the NCBI RefSeq Protein Sequence Records (*i.e.*, those sequences including the designation "NP"). Alternatively, the polynucleotide fragments described in column 2 may refer to assemblages of both cDNA and Genscan-predicted exons brought together by an "exon stitching" algorithm. For example, a polynucleotide sequence identified as FL_*XXXXXX*_N_1_N_2_*YYYYY*_N_3_N_4 represents a "stitched" sequence in which *XXXXXX* is the identification number of the cluster of sequences to which the algorithm was applied, and *YYYYY* is the number of the prediction generated by the algorithm, and $N_{1,2,3...}$, if present, represent specific exons that may have been manually edited during analysis (See Example V). Alternatively, the polynucleotide fragments in column 2 may refer to assemblages of exons brought together by an "exon-stretching" algorithm. For example, a polynucleotide sequence identified as FL*XXXXXX*_g*AAAAA*_g*BBBBB*_1_*N* is a "stretched" sequence, with *XXXXXX* being the Incyte project identification number, g*AAAAA* being the GenBank identification number of the human genomic sequence to which the "exon-stretching" algorithm was applied, g*BBBBB* being the GenBank identification number or NCBI RefSeq identification number of the nearest GenBank protein homolog, and *N* referring to specific exons (See Example V). In instances where a RefSeq sequence was used as a protein homolog for the "exon-stretching" algorithm, a RefSeq identifier (denoted by "NM," "NP," or "NT") may be used in place of the GenBank identifier (*i.e.*, g*BBBBB*).

Alternatively, a prefix identifies component sequences that were hand-edited, predicted from genomic DNA sequences, or derived from a combination of sequence analysis methods. The

following Table lists examples of component sequence prefixes and corresponding sequence analysis methods associated with the prefixes (see Example IV and Example V).

Prefix	Type of analysis and/or examples of programs
GNN, GFG, ENST	Exon prediction from genomic sequences using, for example, GENSCAN (Stanford University, CA, USA) or FGENES (Computer Genomics Group, The Sanger Centre, Cambridge, UK).
GBI	Hand-edited analysis of genomic sequences.
FL	Stitched or stretched genomic sequences (see Example V).
INCY	Full length transcript and exon prediction from mapping of EST sequences to the genome. Genomic location and EST composition data are combined to predict the exons and resulting transcript.

In some cases, Incyte cDNA coverage redundant with the sequence coverage shown in Table 4 was obtained to confirm the final consensus polynucleotide sequence, but the relevant Incyte cDNA identification numbers are not shown.

Table 5 shows the representative cDNA libraries for those full length polynucleotides which were assembled using Incyte cDNA sequences. The representative cDNA library is the Incyte cDNA library which is most frequently represented by the Incyte cDNA sequences which were used to assemble and confirm the above polynucleotides. The tissues and vectors which were used to construct the cDNA libraries shown in Table 5 are described in Table 6.

Table 8 shows single nucleotide polymorphisms (SNPs) found in polynucleotide sequences of the invention, along with allele frequencies in different human populations. Columns 1 and 2 show the polynucleotide sequence identification number (SEQ ID NO:) and the corresponding Incyte project identification number (PID) for polynucleotides of the invention. Column 3 shows the Incyte identification number for the EST in which the SNP was detected (EST ID), and column 4 shows the identification number for the SNP (SNP ID). Column 5 shows the position within the EST sequence at which the SNP is located (EST SNP), and column 6 shows the position of the SNP within the full-length polynucleotide sequence (CB1 SNP). Column 7 shows the allele found in the EST sequence. Columns 8 and 9 show the two alleles found at the SNP site. Column 10 shows the amino acid encoded by the codon including the SNP site, based upon the allele found in the EST. Columns 11-14 show the frequency of allele 1 in four different human populations. An entry of n/d (not detected) indicates that the frequency of allele 1 in the population was too low to be detected, while n/a (not

available) indicates that the allele frequency was not determined for the population.

The invention also encompasses NTRAN variants. Various embodiments of NTRAN variants can have at least about 80%, at least about 90%, or at least about 95% amino acid sequence identity to the NTRAN amino acid sequence, and can contain at least one functional or structural characteristic of NTRAN.

Various embodiments also encompass polynucleotides which encode NTRAN. In a particular embodiment, the invention encompasses a polynucleotide sequence comprising a sequence selected from the group consisting of SEQ ID NO:23-44, which encodes NTRAN. The polynucleotide sequences of SEQ ID NO:23-44, as presented in the Sequence Listing, embrace the equivalent RNA sequences, wherein occurrences of the nitrogenous base thymine are replaced with uracil, and the sugar backbone is composed of ribose instead of deoxyribose.

The invention also encompasses variants of a polynucleotide encoding NTRAN. In particular, such a variant polynucleotide will have at least about 70%, or alternatively at least about 85%, or even at least about 95% polynucleotide sequence identity to a polynucleotide encoding NTRAN. A particular aspect of the invention encompasses a variant of a polynucleotide comprising a sequence selected from the group consisting of SEQ ID NO:23-44 which has at least about 70%, or alternatively at least about 85%, or even at least about 95% polynucleotide sequence identity to a nucleic acid sequence selected from the group consisting of SEQ ID NO:23-44. Any one of the polynucleotide variants described above can encode a polypeptide which contains at least one functional or structural characteristic of NTRAN.

In addition, or in the alternative, a polynucleotide variant of the invention is a splice variant of a polynucleotide encoding NTRAN. A splice variant may have portions which have significant sequence identity to a polynucleotide encoding NTRAN, but will generally have a greater or lesser number of polynucleotides due to additions or deletions of blocks of sequence arising from alternate splicing of exons during mRNA processing. A splice variant may have less than about 70%, or alternatively less than about 60%, or alternatively less than about 50% polynucleotide sequence identity to a polynucleotide encoding NTRAN over its entire length; however, portions of the splice variant will have at least about 70%, or alternatively at least about 85%, or alternatively at least about 95%, or alternatively 100% polynucleotide sequence identity to portions of the polynucleotide encoding NTRAN. For example, a polynucleotide comprising a sequence of SEQ ID NO:33 and a polynucleotide comprising a sequence of SEQ ID NO:34 are splice variants of each other. Any one of the splice variants described above can encode a polypeptide which contains at least one functional

or structural characteristic of NTRAN.

It will be appreciated by those skilled in the art that as a result of the degeneracy of the genetic code, a multitude of polynucleotide sequences encoding NTRAN, some bearing minimal similarity to the polynucleotide sequences of any known and naturally occurring gene, may be produced. Thus, the invention contemplates each and every possible variation of polynucleotide sequence that could be made by selecting combinations based on possible codon choices. These combinations are made in accordance with the standard triplet genetic code as applied to the polynucleotide sequence of naturally occurring NTRAN, and all such variations are to be considered as being specifically disclosed.

Although polynucleotides which encode NTRAN and its variants are generally capable of hybridizing to polynucleotides encoding naturally occurring NTRAN under appropriately selected conditions of stringency, it may be advantageous to produce polynucleotides encoding NTRAN or its derivatives possessing a substantially different codon usage, e.g., inclusion of non-naturally occurring codons. Codons may be selected to increase the rate at which expression of the peptide occurs in a particular prokaryotic or eukaryotic host in accordance with the frequency with which particular codons are utilized by the host. Other reasons for substantially altering the nucleotide sequence encoding NTRAN and its derivatives without altering the encoded amino acid sequences include the production of RNA transcripts having more desirable properties, such as a greater half-life, than transcripts produced from the naturally occurring sequence.

The invention also encompasses production of polynucleotides which encode NTRAN and NTRAN derivatives, or fragments thereof, entirely by synthetic chemistry. After production, the synthetic polynucleotide may be inserted into any of the many available expression vectors and cell systems using reagents well known in the art. Moreover, synthetic chemistry may be used to introduce mutations into a polynucleotide encoding NTRAN or any fragment thereof.

Embodiments of the invention can also include polynucleotides that are capable of hybridizing to the claimed polynucleotides, and, in particular, to those having the sequences shown in SEQ ID NO:23-44 and fragments thereof, under various conditions of stringency (Wahl, G.M. and S.L. Berger (1987) Methods Enzymol. 152:399-407; Kimmel, A.R. (1987) Methods Enzymol. 152:507-511). Hybridization conditions, including annealing and wash conditions, are described in "Definitions."

Methods for DNA sequencing are well known in the art and may be used to practice any of the embodiments of the invention. The methods may employ such enzymes as the Klenow fragment of DNA polymerase I, SEQUENASE (US Biochemical, Cleveland OH), Taq polymerase (Applied

Biosystems), thermostable T7 polymerase (Amersham Biosciences, Piscataway NJ), or combinations of polymerases and proofreading exonucleases such as those found in the ELONGASE amplification system (Invitrogen, Carlsbad CA). Preferably, sequence preparation is automated with machines such as the MICROLAB 2200 liquid transfer system (Hamilton, Reno NV), PTC200 thermal cycler (MJ Research, Watertown MA) and ABI CATALYST 800 thermal cycler (Applied Biosystems). Sequencing is then carried out using either the ABI 373 or 377 DNA sequencing system (Applied Biosystems), the MEGABACE 1000 DNA sequencing system (Amersham Biosciences), or other systems known in the art. The resulting sequences are analyzed using a variety of algorithms which are well known in the art (Ausubel et al., *supra*, ch. 7; Meyers, R.A. (1995) Molecular Biology and Biotechnology, Wiley VCH, New York NY, pp. 856-853).

The nucleic acids encoding NTRAN may be extended utilizing a partial nucleotide sequence and employing various PCR-based methods known in the art to detect upstream sequences, such as promoters and regulatory elements. For example, one method which may be employed, restriction-site PCR, uses universal and nested primers to amplify unknown sequence from genomic DNA within a cloning vector (Sarkar, G. (1993) PCR Methods Applic. 2:318-322). Another method, inverse PCR, uses primers that extend in divergent directions to amplify unknown sequence from a circularized template. The template is derived from restriction fragments comprising a known genomic locus and surrounding sequences (Triglia, T. et al. (1988) Nucleic Acids Res. 16:8186).. A third method, capture PCR, involves PCR amplification of DNA fragments adjacent to known sequences in human and yeast artificial chromosome DNA (Lagerstrom, M. et al. (1991) PCR Methods Applic. 1:111-119). In this method, multiple restriction enzyme digestions and ligations may be used to insert an engineered double-stranded sequence into a region of unknown sequence before performing PCR. Other methods which may be used to retrieve unknown sequences are known in the art (Parker, J.D. et al. (1991) Nucleic Acids Res. 19:3055-3060). Additionally, one may use PCR, nested primers, and PROMOTERFINDER libraries (Clontech, Palo Alto CA) to walk genomic DNA. This procedure avoids the need to screen libraries and is useful in finding intron/exon junctions. For all PCR-based methods, primers may be designed using commercially available software, such as OLIGO 4.06 primer analysis software (National Biosciences, Plymouth MN) or another appropriate program, to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the template at temperatures of about 68°C to 72°C.

When screening for full length cDNAs, it is preferable to use libraries that have been size-selected to include larger cDNAs. In addition, random-primed libraries, which often include

sequences containing the 5' regions of genes, are preferable for situations in which an oligo d(T) library does not yield a full-length cDNA. Genomic libraries may be useful for extension of sequence into 5' non-transcribed regulatory regions.

Capillary electrophoresis systems which are commercially available may be used to analyze the size or confirm the nucleotide sequence of sequencing or PCR products. In particular, capillary sequencing may employ flowable polymers for electrophoretic separation, four different nucleotide-specific, laser-stimulated fluorescent dyes, and a charge coupled device camera for detection of the emitted wavelengths. Output/light intensity may be converted to electrical signal using appropriate software (e.g., GENOTYPER and SEQUENCE NAVIGATOR, Applied Biosystems), and the entire process from loading of samples to computer analysis and electronic data display may be computer controlled. Capillary electrophoresis is especially preferable for sequencing small DNA fragments which may be present in limited amounts in a particular sample.

In another embodiment of the invention, polynucleotides or fragments thereof which encode NTRAN may be cloned in recombinant DNA molecules that direct expression of NTRAN, or fragments or functional equivalents thereof, in appropriate host cells. Due to the inherent degeneracy of the genetic code, other polynucleotides which encode substantially the same or a functionally equivalent polypeptides may be produced and used to express NTRAN.

The polynucleotides of the invention can be engineered using methods generally known in the art in order to alter NTRAN-encoding sequences for a variety of purposes including, but not limited to, modification of the cloning, processing, and/or expression of the gene product. DNA shuffling by random fragmentation and PCR reassembly of gene fragments and synthetic oligonucleotides may be used to engineer the nucleotide sequences. For example, oligonucleotide-mediated site-directed mutagenesis may be used to introduce mutations that create new restriction sites, alter glycosylation patterns, change codon preference, produce splice variants, and so forth.

The nucleotides of the present invention may be subjected to DNA shuffling techniques such as MOLECULARBREEDING (Maxygen Inc., Santa Clara CA; described in U.S. Patent No. 5,837,458; Chang, C.-C. et al. (1999) Nat. Biotechnol. 17:793-797; Christians, F.C. et al. (1999) Nat. Biotechnol. 17:259-264; and Crameri, A. et al. (1996) Nat. Biotechnol. 14:315-319) to alter or improve the biological properties of NTRAN, such as its biological or enzymatic activity or its ability to bind to other molecules or compounds. DNA shuffling is a process by which a library of gene variants is produced using PCR-mediated recombination of gene fragments. The library is then subjected to selection or screening procedures that identify those gene variants with the desired properties. These

preferred variants may then be pooled and further subjected to recursive rounds of DNA shuffling and selection/screening. Thus, genetic diversity is created through "artificial" breeding and rapid molecular evolution. For example, fragments of a single gene containing random point mutations may be recombined, screened, and then reshuffled until the desired properties are optimized. Alternatively, fragments of a given gene may be recombined with fragments of homologous genes in the same gene family, either from the same or different species, thereby maximizing the genetic diversity of multiple naturally occurring genes in a directed and controllable manner.

In another embodiment, polynucleotides encoding NTRAN may be synthesized, in whole or in part, using one or more chemical methods well known in the art (Caruthers, M.H. et al. (1980) Nucleic Acids Symp. Ser. 7:215-223; Horn, T. et al. (1980) Nucleic Acids Symp. Ser. 7:225-232). Alternatively, NTRAN itself or a fragment thereof may be synthesized using chemical methods known in the art. For example, peptide synthesis can be performed using various solution-phase or solid-phase techniques (Creighton, T. (1984) Proteins, Structures and Molecular Properties, WH Freeman, New York NY, pp. 55-60; Roberge, J.Y. et al. (1995) Science 269:202-204). Automated synthesis may be achieved using the ABI 431A peptide synthesizer (Applied Biosystems). Additionally, the amino acid sequence of NTRAN, or any part thereof, may be altered during direct synthesis and/or combined with sequences from other proteins, or any part thereof, to produce a variant polypeptide or a polypeptide having a sequence of a naturally occurring polypeptide.

The peptide may be substantially purified by preparative high performance liquid chromatography (Chiez, R.M. and F.Z. Regnier (1990) Methods Enzymol. 182:392-421). The composition of the synthetic peptides may be confirmed by amino acid analysis or by sequencing (Creighton, *supra*, pp. 28-53).

In order to express a biologically active NTRAN, the polynucleotides encoding NTRAN or derivatives thereof may be inserted into an appropriate expression vector, i.e., a vector which contains the necessary elements for transcriptional and translational control of the inserted coding sequence in a suitable host. These elements include regulatory sequences, such as enhancers, constitutive and inducible promoters, and 5' and 3' untranslated regions in the vector and in polynucleotides encoding NTRAN. Such elements may vary in their strength and specificity. Specific initiation signals may also be used to achieve more efficient translation of polynucleotides encoding NTRAN. Such signals include the ATG initiation codon and adjacent sequences, e.g. the Kozak sequence. In cases where a polynucleotide sequence encoding NTRAN and its initiation codon and upstream regulatory sequences are inserted into the appropriate expression vector, no additional transcriptional or translational control

signals may be needed. However, in cases where only coding sequence, or a fragment thereof, is inserted, exogenous translational control signals including an in-frame ATG initiation codon should be provided by the vector. Exogenous translational elements and initiation codons may be of various origins, both natural and synthetic. The efficiency of expression may be enhanced by the inclusion of enhancers appropriate for the particular host cell system used (Scharf, D. et al. (1994) Results Probl. Cell Differ. 20:125-162).

Methods which are well known to those skilled in the art may be used to construct expression vectors containing polynucleotides encoding NTRAN and appropriate transcriptional and translational control elements. These methods include *in vitro* recombinant DNA techniques, synthetic techniques, and *in vivo* genetic recombination (Sambrook and Russell, *supra*, ch. 1-4, and 8; Ausubel et al., *supra*, ch. 1, 3, and 15).

A variety of expression vector/host systems may be utilized to contain and express polynucleotides encoding NTRAN. These include, but are not limited to, microorganisms such as bacteria transformed with recombinant bacteriophage, plasmid, or cosmid DNA expression vectors; yeast transformed with yeast expression vectors; insect cell systems infected with viral expression vectors (e.g., baculovirus); plant cell systems transformed with viral expression vectors (e.g., cauliflower mosaic virus, CaMV, or tobacco mosaic virus, TMV) or with bacterial expression vectors (e.g., Ti or pBR322 plasmids); or animal cell systems (Sambrook and Russell, *supra*; Ausubel et al., *supra*; Van Heeke, G. and S.M. Schuster (1989) J. Biol. Chem. 264:5503-5509; Engelhard, E.K. et al. (1994) Proc. Natl. Acad. Sci. USA 91:3224-3227; Sandig, V. et al. (1996) Hum. Gene Ther. 7:1937-1945; Takamatsu, N. (1987) EMBO J. 6:307-311; The McGraw Hill Yearbook of Science and Technology (1992) McGraw Hill, New York NY, pp. 191-196; Logan, J. and T. Shenk (1984) Proc. Natl. Acad. Sci. USA 81:3655-3659; Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355). Expression vectors derived from retroviruses, adenoviruses, or herpes or vaccinia viruses, or from various bacterial plasmids, may be used for delivery of polynucleotides to the targeted organ, tissue, or cell population (Di Nicola, M. et al. (1998) Cancer Gen. Ther. 5:350-356; Yu, M. et al. (1993) Proc. Natl. Acad. Sci. USA 90:6340-6344; Buller, R.M. et al. (1985) Nature 317:813-815; McGregor, D.P. et al. (1994) Mol. Immunol. 31:219-226; Verma, I.M. and N. Somia (1997) Nature 389:239-242). The invention is not limited by the host cell employed.

In bacterial systems, a number of cloning and expression vectors may be selected depending upon the use intended for polynucleotides encoding NTRAN. For example, routine cloning, subcloning, and propagation of polynucleotides encoding NTRAN can be achieved using a

multifunctional *E. coli* vector such as PBLUESCRIPT (Stratagene, La Jolla CA) or PSPORT1 plasmid (Invitrogen). Ligation of polynucleotides encoding NTRAN into the vector's multiple cloning site disrupts the *lacZ* gene, allowing a colorimetric screening procedure for identification of transformed bacteria containing recombinant molecules. In addition, these vectors may be useful for *in vitro* transcription, dideoxy sequencing, single strand rescue with helper phage, and creation of nested deletions in the cloned sequence (Van Heeke, G. and S.M. Schuster (1989) J. Biol. Chem. 264:5503-5509). When large quantities of NTRAN are needed, e.g. for the production of antibodies, vectors which direct high level expression of NTRAN may be used. For example, vectors containing the strong, inducible SP6 or T7 bacteriophage promoter may be used.

Yeast expression systems may be used for production of NTRAN. A number of vectors containing constitutive or inducible promoters, such as alpha factor, alcohol oxidase, and PGH promoters, may be used in the yeast *Saccharomyces cerevisiae* or *Pichia pastoris*. In addition, such vectors direct either the secretion or intracellular retention of expressed proteins and enable integration of foreign polynucleotide sequences into the host genome for stable propagation (Ausubel et al., *supra*; Bitter, G.A. et al. (1987) Methods Enzymol. 153:516-544; Scorer, C.A. et al. (1994) Bio/Technology 12:181-184).

Plant systems may also be used for expression of NTRAN. Transcription of polynucleotides encoding NTRAN may be driven by viral promoters, e.g., the 35S and 19S promoters of CaMV used alone or in combination with the omega leader sequence from TMV (Takamatsu, N. (1987) EMBO J. 6:307-311). Alternatively, plant promoters such as the small subunit of RUBISCO or heat shock promoters may be used (Coruzzi, G. et al. (1984) EMBO J. 3:1671-1680; Broglie, R. et al. (1984) Science 224:838-843; Winter, J. et al. (1991) Results Probl. Cell Differ. 17:85-105). These constructs can be introduced into plant cells by direct DNA transformation or pathogen-mediated transfection (The McGraw Hill Yearbook of Science and Technology (1992) McGraw Hill, New York NY, pp. 191-196).

In mammalian cells, a number of viral-based expression systems may be utilized. In cases where an adenovirus is used as an expression vector, polynucleotides encoding NTRAN may be ligated into an adenovirus transcription/translation complex consisting of the late promoter and tripartite leader sequence. Insertion in a non-essential E1 or E3 region of the viral genome may be used to obtain infective virus which expresses NTRAN in host cells (Logan, J. and T. Shenk (1984) Proc. Natl. Acad. Sci. USA 81:3655-3659). In addition, transcription enhancers, such as the Rous sarcoma virus (RSV) enhancer, may be used to increase expression in mammalian host cells. SV40 or EBV-

based vectors may also be used for high-level protein expression.

Human artificial chromosomes (HACs) may also be employed to deliver larger fragments of DNA than can be contained in and expressed from a plasmid. HACs of about 6 kb to 10 Mb are constructed and delivered via conventional delivery methods (liposomes, polycationic amino polymers, or vesicles) for therapeutic purposes (Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355).

For long term production of recombinant proteins in mammalian systems, stable expression of NTRAN in cell lines is preferred. For example, polynucleotides encoding NTRAN can be transformed into cell lines using expression vectors which may contain viral origins of replication and/or endogenous expression elements and a selectable marker gene on the same or on a separate vector. Following the introduction of the vector, cells may be allowed to grow for about 1 to 2 days in enriched media before being switched to selective media. The purpose of the selectable marker is to confer resistance to a selective agent, and its presence allows growth and recovery of cells which successfully express the introduced sequences. Resistant clones of stably transformed cells may be propagated using tissue culture techniques appropriate to the cell type.

Any number of selection systems may be used to recover transformed cell lines. These include, but are not limited to, the herpes simplex virus thymidine kinase and adenine phosphoribosyltransferase genes, for use in *tk*⁻ and *apr*⁻ cells, respectively (Wigler, M. et al. (1977) Cell 11:223-232; Lowy, I. et al. (1980) Cell 22:817-823). Also, antimetabolite, antibiotic, or herbicide resistance can be used as the basis for selection. For example, *dhfr* confers resistance to methotrexate; *neo* confers resistance to the aminoglycosides neomycin and G-418; and *als* and *pat* confer resistance to chlorsulfuron and phosphinotricin acetyltransferase, respectively (Wigler, M. et al. (1980) Proc. Natl. Acad. Sci. USA 77:3567-3570; Colbere-Garapin, F. et al. (1981) J. Mol. Biol. 150:1-14). Additional selectable genes have been described, e.g., *trpB* and *hisD*, which alter cellular requirements for metabolites (Hartman, S.C. and R.C. Mulligan (1988) Proc. Natl. Acad. Sci. USA 85:8047-8051). Visible markers, e.g., anthocyanins, green fluorescent proteins (GFP; Clontech), β-glucuronidase and its substrate β-glucuronide, or luciferase and its substrate luciferin may be used. These markers can be used not only to identify transformants, but also to quantify the amount of transient or stable protein expression attributable to a specific vector system (Rhodes, C.A. (1995) Methods Mol. Biol. 55:121-131).

Although the presence/absence of marker gene expression suggests that the gene of interest is also present, the presence and expression of the gene may need to be confirmed. For example, if the sequence encoding NTRAN is inserted within a marker gene sequence, transformed cells

containing polynucleotides encoding NTRAN can be identified by the absence of marker gene function. Alternatively, a marker gene can be placed in tandem with a sequence encoding NTRAN under the control of a single promoter. Expression of the marker gene in response to induction or selection usually indicates expression of the tandem gene as well.

In general, host cells that contain the polynucleotide encoding NTRAN and that express NTRAN may be identified by a variety of procedures known to those of skill in the art. These procedures include, but are not limited to, DNA-DNA or DNA-RNA hybridizations, PCR amplification, and protein bioassay or immunoassay techniques which include membrane, solution, or chip based technologies for the detection and/or quantification of nucleic acid or protein sequences.

Immunological methods for detecting and measuring the expression of NTRAN using either specific polyclonal or monoclonal antibodies are known in the art. Examples of such techniques include enzyme-linked immunosorbent assays (ELISAs), radioimmunoassays (RIAs), and fluorescence activated cell sorting (FACS). A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering epitopes on NTRAN is preferred, but a competitive binding assay may be employed. These and other assays are well known in the art (Hampton, R. et al. (1990) Serological Methods, a Laboratory Manual, APS Press, St. Paul MN, Sect. IV; Coligan, J.E. et al. (1997) Current Protocols in Immunology, Greene Pub. Associates and Wiley-Interscience, New York NY; Pound, J.D. (1998) Immunochemical Protocols, Humana Press, Totowa NJ).

A wide variety of labels and conjugation techniques are known by those skilled in the art and may be used in various nucleic acid and amino acid assays. Means for producing labeled hybridization or PCR probes for detecting sequences related to polynucleotides encoding NTRAN include oligolabeling, nick translation, end-labeling, or PCR amplification using a labeled nucleotide. Alternatively, polynucleotides encoding NTRAN, or any fragments thereof, may be cloned into a vector for the production of an mRNA probe. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes *in vitro* by addition of an appropriate RNA polymerase such as T7, T3, or SP6 and labeled nucleotides. These procedures may be conducted using a variety of commercially available kits, such as those provided by Amersham Biosciences, Promega (Madison WI), and US Biochemical. Suitable reporter molecules or labels which may be used for ease of detection include radionuclides, enzymes, fluorescent, chemiluminescent, or chromogenic agents, as well as substrates, cofactors, inhibitors, magnetic particles, and the like.

Host cells transformed with polynucleotides encoding NTRAN may be cultured under

conditions suitable for the expression and recovery of the protein from cell culture. The protein produced by a transformed cell may be secreted or retained intracellularly depending on the sequence and/or the vector used. As will be understood by those of skill in the art, expression vectors containing polynucleotides which encode NTRAN may be designed to contain signal sequences which direct secretion of NTRAN through a prokaryotic or eukaryotic cell membrane.

In addition, a host cell strain may be chosen for its ability to modulate expression of the inserted polynucleotides or to process the expressed protein in the desired fashion. Such modifications of the polypeptide include, but are not limited to, acetylation, carboxylation, glycosylation, phosphorylation, lipidation, and acylation. Post-translational processing which cleaves a "prepro" or "pro" form of the protein may also be used to specify protein targeting, folding, and/or activity. Different host cells which have specific cellular machinery and characteristic mechanisms for post-translational activities (e.g., CHO, HeLa, MDCK, HEK293, and WI38) are available from the American Type Culture Collection (ATCC, Manassas VA) and may be chosen to ensure the correct modification and processing of the foreign protein.

In another embodiment of the invention, natural, modified, or recombinant polynucleotides encoding NTRAN may be ligated to a heterologous sequence resulting in translation of a fusion protein in any of the aforementioned host systems. For example, a chimeric NTRAN protein containing a heterologous moiety that can be recognized by a commercially available antibody may facilitate the screening of peptide libraries for inhibitors of NTRAN activity. Heterologous protein and peptide moieties may also facilitate purification of fusion proteins using commercially available affinity matrices. Such moieties include, but are not limited to, glutathione S-transferase (GST), maltose binding protein (MBP), thioredoxin (Trx), calmodulin binding peptide (CBP), 6-His, FLAG, *c-myc*, and hemagglutinin (HA). GST, MBP, Trx, CBP, and 6-His enable purification of their cognate fusion proteins on immobilized glutathione, maltose, phenylarsine oxide, calmodulin, and metal-chelate resins, respectively. FLAG, *c-myc*, and hemagglutinin (HA) enable immunoaffinity purification of fusion proteins using commercially available monoclonal and polyclonal antibodies that specifically recognize these epitope tags. A fusion protein may also be engineered to contain a proteolytic cleavage site located between the NTRAN encoding sequence and the heterologous protein sequence, so that NTRAN may be cleaved away from the heterologous moiety following purification. Methods for fusion protein expression and purification are discussed in Ausubel et al. (*supra*, ch. 10 and 16). A variety of commercially available kits may also be used to facilitate expression and purification of fusion proteins.

In another embodiment, synthesis of radiolabeled NTRAN may be achieved *in vitro* using the TNT rabbit reticulocyte lysate or wheat germ extract system (Promega). These systems couple transcription and translation of protein-coding sequences operably associated with the T7, T3, or SP6 promoters. Translation takes place in the presence of a radiolabeled amino acid precursor, for example, ^{35}S-methionine.

NTRAN, fragments of NTRAN, or variants of NTRAN may be used to screen for compounds that specifically bind to NTRAN. One or more test compounds may be screened for specific binding to NTRAN. In various embodiments, 1, 2, 3, 4, 5, 10, 20, 50, 100, or 200 test compounds can be screened for specific binding to NTRAN. Examples of test compounds can include antibodies, anticalins, oligonucleotides, proteins (e.g., ligands or receptors), or small molecules.

In related embodiments, variants of NTRAN can be used to screen for binding of test compounds, such as antibodies, to NTRAN, a variant of NTRAN, or a combination of NTRAN and/or one or more variants NTRAN. In an embodiment, a variant of NTRAN can be used to screen for compounds that bind to a variant of NTRAN, but not to NTRAN having the exact sequence of a sequence of SEQ ID NO:1-22. NTRAN variants used to perform such screening can have a range of about 50% to about 99% sequence identity to NTRAN, with various embodiments having 60%, 70%, 75%, 80%, 85%, 90%, and 95% sequence identity.

In an embodiment, a compound identified in a screen for specific binding to NTRAN can be closely related to the natural ligand of NTRAN, e.g., a ligand or fragment thereof, a natural substrate, a structural or functional mimetic, or a natural binding partner (Coligan, J.E. et al. (1991) Current Protocols in Immunology 1(2):Chapter 5). In another embodiment, the compound thus identified can be a natural ligand of a receptor NTRAN (Howard, A.D. et al. (2001) Trends Pharmacol. Sci.22:132-140; Wise, A. et al. (2002) Drug Discovery Today 7:235-246).

In other embodiments, a compound identified in a screen for specific binding to NTRAN can be closely related to the natural receptor to which NTRAN binds, at least a fragment of the receptor, or a fragment of the receptor including all or a portion of the ligand binding site or binding pocket. For example, the compound may be a receptor for NTRAN which is capable of propagating a signal, or a decoy receptor for NTRAN which is not capable of propagating a signal (Ashkenazi, A. and V.M. Divit (1999) Curr. Opin. Cell Biol. 11:255-260; Mantovani, A. et al. (2001) Trends Immunol. 22:328-336). The compound can be rationally designed using known techniques. Examples of such techniques include those used to construct the compound etanercept (ENBREL; Amgen Inc., Thousand Oaks CA), which is efficacious for treating rheumatoid arthritis in humans. Etanercept is

an engineered p75 tumor necrosis factor (TNF) receptor dimer linked to the Fc portion of human IgG_1 (Taylor, P.C. et al. (2001) Curr. Opin. Immunol. 13:611-616).

In one embodiment, two or more antibodies having similar or, alternatively, different specificities can be screened for specific binding to NTRAN, fragments of NTRAN, or variants of NTRAN. The binding specificity of the antibodies thus screened can thereby be selected to identify particular fragments or variants of NTRAN. In one embodiment, an antibody can be selected such that its binding specificity allows for preferential identification of specific fragments or variants of NTRAN. In another embodiment, an antibody can be selected such that its binding specificity allows for preferential diagnosis of a specific disease or condition having increased, decreased, or otherwise abnormal production of NTRAN.

In an embodiment, anticalins can be screened for specific binding to NTRAN, fragments of NTRAN, or variants of NTRAN. Anticalins are ligand-binding proteins that have been constructed based on a lipocalin scaffold (Weiss, G.A. and H.B. Lowman (2000) Chem. Biol. 7:R177-R184; Skerra, A. (2001) J. Biotechnol. 74:257-275). The protein architecture of lipocalins can include a beta-barrel having eight antiparallel beta-strands, which supports four loops at its open end. These loops form the natural ligand-binding site of the lipocalins, a site which can be re-engineered *in vitro* by amino acid substitutions to impart novel binding specificities. The amino acid substitutions can be made using methods known in the art or described herein, and can include conservative substitutions (e.g., substitutions that do not alter binding specificity) or substitutions that modestly, moderately, or significantly alter binding specificity.

In one embodiment, screening for compounds which specifically bind to, stimulate, or inhibit NTRAN involves producing appropriate cells which express NTRAN, either as a secreted protein or on the cell membrane. Preferred cells can include cells from mammals, yeast, *Drosophila*, or *E. coli*. Cells expressing NTRAN or cell membrane fractions which contain NTRAN are then contacted with a test compound and binding, stimulation, or inhibition of activity of either NTRAN or the compound is analyzed.

An assay may simply test binding of a test compound to the polypeptide, wherein binding is detected by a fluorophore, radioisotope, enzyme conjugate, or other detectable label. For example, the assay may comprise the steps of combining at least one test compound with NTRAN, either in solution or affixed to a solid support, and detecting the binding of NTRAN to the compound. Alternatively, the assay may detect or measure binding of a test compound in the presence of a labeled competitor. Additionally, the assay may be carried out using cell-free preparations, chemical

libraries, or natural product mixtures, and the test compound(s) may be free in solution or affixed to a solid support.

An assay can be used to assess the ability of a compound to bind to its natural ligand and/or to inhibit the binding of its natural ligand to its natural receptors. Examples of such assays include radio-labeling assays such as those described in U.S. Patent No. 5,914,236 and U.S. Patent No. 6,372,724. In a related embodiment, one or more amino acid substitutions can be introduced into a polypeptide compound (such as a receptor) to improve or alter its ability to bind to its natural ligands (Matthews, D.J. and J.A. Wells. (1994) Chem. Biol. 1:25-30). In another related embodiment, one or more amino acid substitutions can be introduced into a polypeptide compound (such as a ligand) to improve or alter its ability to bind to its natural receptors (Cunningham, B.C. and J.A. Wells (1991) Proc. Natl. Acad. Sci. USA 88:3407-3411; Lowman, H.B. et al. (1991) J. Biol. Chem. 266:10982-10988).

NTRAN, fragments of NTRAN, or variants of NTRAN may be used to screen for compounds that modulate the activity of NTRAN. Such compounds may include agonists, antagonists, or partial or inverse agonists. In one embodiment, an assay is performed under conditions permissive for NTRAN activity, wherein NTRAN is combined with at least one test compound, and the activity of NTRAN in the presence of a test compound is compared with the activity of NTRAN in the absence of the test compound. A change in the activity of NTRAN in the presence of the test compound is indicative of a compound that modulates the activity of NTRAN. Alternatively, a test compound is combined with an *in vitro* or cell-free system comprising NTRAN under conditions suitable for NTRAN activity, and the assay is performed. In either of these assays, a test compound which modulates the activity of NTRAN may do so indirectly and need not come in direct contact with the test compound. At least one and up to a plurality of test compounds may be screened.

In another embodiment, polynucleotides encoding NTRAN or their mammalian homologs may be "knocked out" in an animal model system using homologous recombination in embryonic stem (ES) cells. Such techniques are well known in the art and are useful for the generation of animal models of human disease (see, e.g., U.S. Patent No. 5,175,383 and U.S. Patent No. 5,767,337). For example, mouse ES cells, such as the mouse 129/SvJ cell line, are derived from the early mouse embryo and grown in culture. The ES cells are transformed with a vector containing the gene of interest disrupted by a marker gene, e.g., the neomycin phosphotransferase gene (*neo*; Capecchi, M.R. (1989) Science 244:1288-1292). The vector integrates into the corresponding region of the host genome by homologous recombination. Alternatively, homologous recombination takes place using the Cre-loxP system to knockout a gene of interest in a tissue- or developmental stage-specific manner (Marth, J.D.

(1996) Clin. Invest. 97:1999-2002; Wagner, K.U. et al. (1997) Nucleic Acids Res. 25:4323-4330). Transformed ES cells are identified and microinjected into mouse cell blastocysts such as those from the C57BL/6 mouse strain. The blastocysts are surgically transferred to pseudopregnant dams, and the resulting chimeric progeny are genotyped and bred to produce heterozygous or homozygous strains. Transgenic animals thus generated may be tested with potential therapeutic or toxic agents.

Polynucleotides encoding NTRAN may also be manipulated *in vitro* in ES cells derived from human blastocysts. Human ES cells have the potential to differentiate into at least eight separate cell lineages including endoderm, mesoderm, and ectodermal cell types. These cell lineages differentiate into, for example, neural cells, hematopoietic lineages, and cardiomyocytes (Thomson, J.A. et al. (1998) Science 282:1145-1147).

Polynucleotides encoding NTRAN can also be used to create "knockin" humanized animals (pigs) or transgenic animals (mice or rats) to model human disease. With knockin technology, a region of a polynucleotide encoding NTRAN is injected into animal ES cells, and the injected sequence integrates into the animal cell genome. Transformed cells are injected into blastulae, and the blastulae are implanted as described above. Transgenic progeny or inbred lines are studied and treated with potential pharmaceutical agents to obtain information on treatment of a human disease. Alternatively, a mammal inbred to overexpress NTRAN, e.g., by secreting NTRAN in its milk, may also serve as a convenient source of that protein (Janne, J. et al. (1998) Biotechnol. Annu. Rev. 4:55-74).

THERAPEUTICS

Chemical and structural similarity, e.g., in the context of sequences and motifs, exists between regions of NTRAN and neurotransmission-associated proteins. In addition, examples of tissues expressing NTRAN can be found in Table 6 and can also be found in Example XI. Therefore, NTRAN appears to play a role in autoimmune/inflammatory, cardiovascular, neurological, developmental, cell proliferative, transport, psychiatric, metabolic, and endocrine disorders. In the treatment of disorders associated with increased NTRAN expression or activity, it is desirable to decrease the expression or activity of NTRAN. In the treatment of disorders associated with decreased NTRAN expression or activity, it is desirable to increase the expression or activity of NTRAN.

Therefore, in one embodiment, NTRAN or a fragment or derivative thereof may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of NTRAN. Examples of such disorders include, but are not limited to, an autoimmune/inflammatory disorder such as acquired immunodeficiency syndrome (AIDS), Addison's

disease, adult respiratory distress syndrome, allergies, ankylosing spondylitis, amyloidosis, anemia, asthma, atherosclerosis, autoimmune hemolytic anemia, autoimmune thyroiditis, autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED), bronchitis, cholecystitis, contact dermatitis, Crohn's disease, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, episodic lymphopenia with lymphocytotoxins, erythroblastosis fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis, hypereosinophilia, irritable bowel syndrome, multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis, thromb osteoarthritis, osteoporosis, pancreatitis, polymyositis, psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjögren's ocytopenic purpura, ulcerative colitis, uveitis, Werner syndrome, complications of cancer, hemodialysis, and extracorporeal circulation, viral, bacterial, fungal, parasitic, protozoal, and helminthic infections, and trauma; a cardiovascular disorder such as congestive heart failure, ischemic heart disease, angina pectoris, myocardial infarction, hypertensive heart disease, degenerative valvular heart disease, calcific aortic valve stenosis, congenitally bicuspid aortic valve, mitral annular calcification, mitral valve prolapse, rheumatic fever and rheumatic heart disease, infective endocarditis, nonbacterial thrombotic endocarditis, endocarditis of systemic lupus erythematosus, carcinoid heart disease, cardiomyopathy, myocarditis, pericarditis, neoplastic heart disease, congenital heart disease, and complications of cardiac transplantation, arteriovenous fistula, atherosclerosis, hypertension, vasculitis, Raynaud's disease, aneurysms, arterial dissections, varicose veins, thrombophlebitis and phlebothrombosis, vascular tumors, and complications of thrombolysis, balloon angioplasty, vascular replacement, and coronary artery bypass graft surgery; a neurological disorder such as epilepsy, ischemic cerebrovascular disease, stroke, cerebral neoplasms, Pick's disease, Huntington's disease, dementia, Parkinson's disease and other extrapyramidal disorders, amyotrophic lateral sclerosis and other motor neuron disorders, progressive neural muscular atrophy, retinitis pigmentosa, hereditary ataxias, multiple sclerosis and other demyelinating diseases, bacterial and viral meningitis, brain abscess, subdural empyema, epidural abscess, suppurative intracranial thrombophlebitis, myelitis and radiculitis, viral central nervous system disease, prion diseases including kuru, Creutzfeldt-Jakob disease, and Gerstmann-Straussler-Scheinker syndrome, fatal familial insomnia, nutritional and metabolic diseases of the nervous system, neurofibromatosis, tuberous sclerosis, cerebelloretinal hemangioblastomatosis, encephalotrigeminal syndrome, mental retardation and other developmental disorders of the central nervous system including Down syndrome, cerebral palsy, neuroskeletal disorders, autonomic nervous system disorders, cranial nerve disorders, spinal cord

diseases, muscular dystrophy and other neuromuscular disorders, peripheral nervous system disorders, dermatomyositis and polymyositis, inherited, metabolic, endocrine, and toxic myopathies, myasthenia gravis, periodic paralysis, mental disorders including mood, anxiety, and schizophrenic disorders, seasonal affective disorder (SAD), akathesia, amnesia, catatonia, diabetic neuropathy, tardive dyskinesia, dystonias, paranoid psychoses, postherpetic neuralgia, Tourette's disorder, progressive supranuclear palsy, corticobasal degeneration, and familial frontotemporal dementia; a developmental disorder such as renal tubular acidosis, anemia, Cushing's syndrome, achondroplastic dwarfism, Duchenne and Becker muscular dystrophy, epilepsy, gonadal dysgenesis, WAGR syndrome (Wilms' tumor, aniridia, genitourinary abnormalities, and mental retardation), Smith-Magenis syndrome, myelodysplastic syndrome, hereditary mucoepithelial dysplasia, hereditary keratodermas, hereditary neuropathies such as Charcot-Marie-Tooth disease and neurofibromatosis, hypothyroidism, hydrocephalus, seizure disorders such as Syndenham's chorea and cerebral palsy, spina bifida, anencephaly, craniorachischisis, congenital glaucoma, cataract, and sensorineural hearing loss; a cell proliferative disorder such as actinic keratosis, arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia, and cancers including adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, cancers of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, colon, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus and a cancer such as adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, cancers of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus; a transport disorder such as akinesia, amyotrophic lateral sclerosis, ataxia telangiectasia, cystic fibrosis, Becker's muscular dystrophy, Bell's palsy, Charcot-Marie Tooth disease, diabetes mellitus, diabetes insipidus, diabetic neuropathy, Duchenne muscular dystrophy, hyperkalemic periodic paralysis, normokalemic periodic paralysis, Parkinson's disease, malignant hyperthermia, multidrug resistance, myasthenia gravis, myotonic dystrophy, catatonia, tardive dyskinesia, dystonias, peripheral neuropathy, cerebral neoplasms, prostate cancer, cardiac disorders associated with transport, e.g., angina, bradyarrythmia, tachyarrythmia, hypertension, Long QT syndrome, myocarditis, cardiomyopathy, nemaline myopathy, centronuclear myopathy, lipid myopathy, mitochondrial myopathy, thyrotoxic myopathy, ethanol

myopathy, dermatomyositis, inclusion body myositis, infectious myositis, polymyositis, neurological disorders associated with transport, e.g., Alzheimer's disease, amnesia, bipolar disorder, dementia, depression, epilepsy, Tourette's disorder, paranoid psychoses, and schizophrenia, and other disorders associated with transport, e.g., neurofibromatosis, postherpetic neuralgia, trigeminal neuropathy, sarcoidosis, sickle cell anemia, Wilson's disease, cataracts, infertility, pulmonary artery stenosis, sensorineural autosomal deafness, hyperglycemia, hypoglycemia, Grave's disease, goiter, Cushing's disease, Addison's disease, glucose-galactose malabsorption syndrome, hypercholesterolemia, adrenoleukodystrophy, Zellweger syndrome, Menkes disease, occipital horn syndrome, von Gierke disease, cystinuria, iminoglycinuria, Hartup disease, and Fanconi disease; a psychiatric disorder such as acute stress disorder, alcohol dependence, amphetamine dependence, anorexia nervosa, antisocial personality disorder, attention-deficit hyperactivity disorder, autistic disorder, anxiety, avoidant personality disorder, bipolar disorder, borderline personality disorder, brief psychotic disorder, bulimia nervosa, cannabis dependence, cocaine dependence, conduct disorder, cyclothymic disorder, delirium, delusional disorder, dementia, dependent personality disorder, depression, dysthymic disorder, hallucinogen dependence, histrionic personality disorder, inhalant dependence, manic depression, multi-infarct dementia, narcissistic personality disorder, nicotine dependence, obsessive-compulsive disorder, opioid dependence, oppositional defiant disorder, panic disorder, paranoid personality disorder, phencyclidine dependence, phobia, posttraumatic stress disorder, schizoaffective disorder, schizoid personality disorder, schizophrenia, sedative dependence, separation anxiety disorder, and sleep disorder; a metabolic disorder such as Addison's disease, cerebrotendinous xanthomatosis, congenital adrenal hyperplasia, coumarin resistance, cystic fibrosis, fatty hepatocirrhosis, fructose-1,6-diphosphatase deficiency, galactosemia, goiter, glucagonoma, glycogen storage diseases, hereditary fructose intolerance, hyperadrenalism, hypoadrenalism, hyperparathyroidism, hypoparathyroidism, hypercholesterolemia, hyperthyroidism, hypoglycemia, hypothyroidism, hyperlipidemia, hyperlipemia, lipid myopathies, lipodystrophies, lysosomal storage diseases, mannosidosis, neuraminidase deficiency, obesity, osteoporosis, phenylketonuria, pseudovitamin D-deficiency rickets, disorders of carbohydrate metabolism such as congenital type II dyserythropoietic anemia, diabetes, insulin-dependent diabetes mellitus, non-insulin-dependent diabetes mellitus, galactose epimerase deficiency, glycogen storage diseases, lysosomal storage diseases, fructosuria, pentosuria, and inherited abnormalities of pyruvate metabolism, disorders of lipid metabolism such as fatty liver, cholestasis, primary biliary cirrhosis, carnitine deficiency, carnitine palmitoyltransferase deficiency, myoadenylate deaminase deficiency, hypertriglyceridemia, lipid storage disorders such

Fabry's disease, Gaucher's disease, Niemann-Pick's disease, metachromatic leukodystrophy, adrenoleukodystrophy, GM_2 gangliosidosis, and ceroid lipofuscinosis, abetalipoproteinemia, Tangier disease, hyperlipoproteinemia, lipodystrophy, lipomatoses, acute panniculitis, disseminated fat necrosis, adiposis dolorosa, lipoid adrenal hyperplasia, minimal change disease, lipomas, atherosclerosis, hypercholesterolemia, hypercholesterolemia with hypertriglyceridemia, primary hypoalphalipoproteinemia, hypothyroidism, renal disease, liver disease, lecithin:cholesterol acyltransferase deficiency, cerebrotendinous xanthomatosis, sitosterolemia, hypocholesterolemia, Tay-Sachs disease, Sandhoff's disease, hyperlipidemia, hyperlipemia, and lipid myopathies, and disorders of copper metabolism such as Menke's disease, Wilson's disease, and Ehlers-Danlos syndrome type IX diabetes; and an endocrine disorder such as a disorder of the hypothalamus and/or pituitary resulting from lesions such as a primary brain tumor, adenoma, infarction associated with pregnancy, hypophysectomy, aneurysm, vascular malformation, thrombosis, infection, immunological disorder, and complication due to head trauma, a disorder associated with hypopituitarism including hypogonadism, Sheehan syndrome, diabetes insipidus, Kallman's disease, Hand-Schuller-Christian disease, Letterer-Siwe disease, sarcoidosis, empty sella syndrome, and dwarfism, a disorder associated with hyperpituitarism including acromegaly, giantism, and syndrome of inappropriate antidiuretic hormone (ADH) secretion (SIADH) often caused by benign adenoma, a disorder associated with hypothyroidism including goiter, myxedema, acute thyroiditis associated with bacterial infection, subacute thyroiditis associated with viral infection, autoimmune thyroiditis (Hashimoto's disease), and cretinism, a disorder associated with hyperthyroidism including thyrotoxicosis and its various forms, Grave's disease, pretibial myxedema, toxic multinodular goiter, thyroid carcinoma, and Plummer's disease, a disorder associated with hyperparathyroidism including Conn disease (chronic hypercalemia), a pancreatic disorder such as Type I or Type II diabetes mellitus and associated complications, a disorder associated with the adrenals such as hyperplasia, carcinoma, or adenoma of the adrenal cortex, hypertension associated with alkalosis, amyloidosis, hypokalemia, Cushing's disease, Liddle's syndrome, and Arnold-Healy-Gordon syndrome, pheochromocytoma tumors, and Addison's disease, a disorder associated with gonadal steroid hormones such as: in women, abnormal prolactin production, infertility, endometriosis, perturbation of the menstrual cycle, polycystic ovarian disease, hyperprolactinemia, isolated gonadotropin deficiency, amenorrhea, galactorrhea, hermaphroditism, hirsutism and virilization, breast cancer, and, in post-menopausal women, osteoporosis, and, in men, Leydig cell deficiency, male climacteric phase, and germinal cell aplasia, a hypergonadal disorder associated with Leydig cell tumors, androgen resistance associated with

absence of androgen receptors, syndrome of 5 α-reductase, and gynecomastia.

In another embodiment, a vector capable of expressing NTRAN or a fragment or derivative thereof may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of NTRAN including, but not limited to, those described above.

In a further embodiment, a composition comprising a substantially purified NTRAN in conjunction with a suitable pharmaceutical carrier may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of NTRAN including, but not limited to, those provided above.

In still another embodiment, an agonist which modulates the activity of NTRAN may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of NTRAN including, but not limited to, those listed above.

In a further embodiment, an antagonist of NTRAN may be administered to a subject to treat or prevent a disorder associated with increased expression or activity of NTRAN. Examples of such disorders include, but are not limited to, those autoimmune/inflammatory, cardiovascular, neurological, developmental, cell proliferative, transport, psychiatric, metabolic, and endocrine disorders described above. In one aspect, an antibody which specifically binds NTRAN may be used directly as an antagonist or indirectly as a targeting or delivery mechanism for bringing a pharmaceutical agent to cells or tissues which express NTRAN.

In an additional embodiment, a vector expressing the complement of the polynucleotide encoding NTRAN may be administered to a subject to treat or prevent a disorder associated with increased expression or activity of NTRAN including, but not limited to, those described above.

In other embodiments, any protein, agonist, antagonist, antibody, complementary sequence, or vector embodiments may be administered in combination with other appropriate therapeutic agents. Selection of the appropriate agents for use in combination therapy may be made by one of ordinary skill in the art, according to conventional pharmaceutical principles. The combination of therapeutic agents may act synergistically to effect the treatment or prevention of the various disorders described above. Using this approach, one may be able to achieve therapeutic efficacy with lower dosages of each agent, thus reducing the potential for adverse side effects.

An antagonist of NTRAN may be produced using methods which are generally known in the art. In particular, purified NTRAN may be used to produce antibodies or to screen libraries of pharmaceutical agents to identify those which specifically bind NTRAN. Antibodies to NTRAN may also be generated using methods that are well known in the art. Such antibodies may include, but are

not limited to, polyclonal, monoclonal, chimeric, and single chain antibodies, Fab fragments, and fragments produced by a Fab expression library. In an embodiment, neutralizing antibodies (i.e., those which inhibit dimer formation) can be used therapeutically. Single chain antibodies (e.g., from camels or llamas) may be potent enzyme inhibitors and may have application in the design of peptide mimetics, and in the development of immuno-adsorbents and biosensors (Muyldermans, S. (2001) J. Biotechnol. 74:277-302).

For the production of antibodies, various hosts including goats, rabbits, rats, mice, camels, dromedaries, llamas, humans, and others may be immunized by injection with NTRAN or with any fragment or oligopeptide thereof which has immunogenic properties. Depending on the host species, various adjuvants may be used to increase immunological response. Such adjuvants include, but are not limited to, Freund's, mineral gels such as aluminum hydroxide, and surface active substances such as lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, KLH, and dinitrophenol. Among adjuvants used in humans, BCG (bacilli Calmette-Guerin) and *Corynebacterium parvum* are especially preferable.

It is preferred that the oligopeptides, peptides, or fragments used to induce antibodies to NTRAN have an amino acid sequence consisting of at least about 5 amino acids, and generally will consist of at least about 10 amino acids. It is also preferable that these oligopeptides, peptides, or fragments are substantially identical to a portion of the amino acid sequence of the natural protein. Short stretches of NTRAN amino acids may be fused with those of another protein, such as KLH, and antibodies to the chimeric molecule may be produced.

Monoclonal antibodies to NTRAN may be prepared using any technique which provides for the production of antibody molecules by continuous cell lines in culture. These include, but are not limited to, the hybridoma technique, the human B-cell hybridoma technique, and the EBV-hybridoma technique (Kohler, G. et al. (1975) Nature 256:495-497; Kozbor, D. et al. (1985) J. Immunol. Methods 81:31-42; Cote, R.J. et al. (1983) Proc. Natl. Acad. Sci. USA 80:2026-2030; Cole, S.P. et al. (1984) Mol. Cell Biol. 62:109-120).

In addition, techniques developed for the production of "chimeric antibodies," such as the splicing of mouse antibody genes to human antibody genes to obtain a molecule with appropriate antigen specificity and biological activity, can be used (Morrison, S.L. et al. (1984) Proc. Natl. Acad. Sci. USA 81:6851-6855; Neuberger, M.S. et al. (1984) Nature 312:604-608; Takeda, S. et al. (1985) Nature 314:452-454). Alternatively, techniques described for the production of single chain antibodies may be adapted, using methods known in the art, to produce NTRAN-specific single chain antibodies.

Antibodies with related specificity, but of distinct idiotypic composition, may be generated by chain shuffling from random combinatorial immunoglobulin libraries (Burton, D.R. (1991) Proc. Natl. Acad. Sci. USA 88:10134-10137).

Antibodies may also be produced by inducing *in vivo* production in the lymphocyte population or by screening immunoglobulin libraries or panels of highly specific binding reagents as disclosed in the literature (Orlandi, R. et al. (1989) Proc. Natl. Acad. Sci. USA 86:3833-3837; Winter, G. et al. (1991) Nature 349:293-299).

Antibody fragments which contain specific binding sites for NTRAN may also be generated. For example, such fragments include, but are not limited to, $F(ab')_2$ fragments produced by pepsin digestion of the antibody molecule and Fab fragments generated by reducing the disulfide bridges of the F(ab')2 fragments. Alternatively, Fab expression libraries may be constructed to allow rapid and easy identification of monoclonal Fab fragments with the desired specificity (Huse, W.D. et al. (1989) Science 246:1275-1281).

Various immunoassays may be used for screening to identify antibodies having the desired specificity. Numerous protocols for competitive binding or immunoradiometric assays using either polyclonal or monoclonal antibodies with established specificities are well known in the art. Such immunoassays typically involve the measurement of complex formation between NTRAN and its specific antibody. A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering NTRAN epitopes is generally used, but a competitive binding assay may also be employed (Pound, *supra*).

Various methods such as Scatchard analysis in conjunction with radioimmunoassay techniques may be used to assess the affinity of antibodies for NTRAN. Affinity is expressed as an association constant, K_a, which is defined as the molar concentration of NTRAN-antibody complex divided by the molar concentrations of free antigen and free antibody under equilibrium conditions. The K_a determined for a preparation of polyclonal antibodies, which are heterogeneous in their affinities for multiple NTRAN epitopes, represents the average affinity, or avidity, of the antibodies for NTRAN. The K_a determined for a preparation of monoclonal antibodies, which are monospecific for a particular NTRAN epitope, represents a true measure of affinity. High-affinity antibody preparations with K_a ranging from about 10^9 to 10^{12} L/mole are preferred for use in immunoassays in which the NTRAN-antibody complex must withstand rigorous manipulations. Low-affinity antibody preparations with K_a ranging from about 10^6 to 10^7 L/mole are preferred for use in immunopurification and similar procedures which ultimately require dissociation of NTRAN, preferably in active form, from the

antibody (Catty, D. (1988) Antibodies, Volume I: A Practical Approach, IRL Press, Washington DC; Liddell, J.E. and A. Cryer (1991) A Practical Guide to Monoclonal Antibodies, John Wiley & Sons, New York NY).

The titer and avidity of polyclonal antibody preparations may be further evaluated to determine the quality and suitability of such preparations for certain downstream applications. For example, a polyclonal antibody preparation containing at least 1-2 mg specific antibody/ml, preferably 5-10 mg specific antibody/ml, is generally employed in procedures requiring precipitation of NTRAN-antibody complexes. Procedures for evaluating antibody specificity, titer, and avidity, and guidelines for antibody quality and usage in various applications, are generally available (Catty, *supra*; Coligan et al., *supra*).

In another embodiment of the invention, polynucleotides encoding NTRAN, or any fragment or complement thereof, may be used for therapeutic purposes. In one aspect, modifications of gene expression can be achieved by designing complementary sequences or antisense molecules (DNA, RNA, PNA, or modified oligonucleotides) to the coding or regulatory regions of the gene encoding NTRAN. Such technology is well known in the art, and antisense oligonucleotides or larger fragments can be designed from various locations along the coding or control regions of sequences encoding NTRAN (Agrawal, S., ed. (1996) Antisense Therapeutics, Humana Press, Totawa NJ).

In therapeutic use, any gene delivery system suitable for introduction of the antisense sequences into appropriate target cells can be used. Antisense sequences can be delivered intracellularly in the form of an expression plasmid which, upon transcription, produces a sequence complementary to at least a portion of the cellular sequence encoding the target protein (Slater, J.E. et al. (1998) J. Allergy Clin. Immunol. 102:469-475; Scanlon, K.J. et al. (1995) 9:1288-1296). Antisense sequences can also be introduced intracellularly through the use of viral vectors, such as retrovirus and adeno-associated virus vectors (Miller, A.D. (1990) Blood 76:271; Ausubel et al., *supra*; Uckert, W. and W. Walther (1994) Pharmacol. Ther. 63:323-347). Other gene delivery mechanisms include liposome-derived systems, artificial viral envelopes, and other systems known in the art (Rossi, J.J. (1995) Br. Med. Bull. 51:217-225; Boado, R.J. et al. (1998) J. Pharm. Sci. 87:1308-1315; Morris, M.C. et al. (1997) Nucleic Acids Res. 25:2730-2736).

In another embodiment of the invention, polynucleotides encoding NTRAN may be used for somatic or germline gene therapy. Gene therapy may be performed to (i) correct a genetic deficiency (e.g., in the cases of severe combined immunodeficiency (SCID)-X1 disease characterized by X-linked inheritance (Cavazzana-Calvo, M. et al. (2000) Science 288:669-672), severe combined

immunodeficiency syndrome associated with an inherited adenosine deaminase (ADA) deficiency (Blaese, R.M. et al. (1995) Science 270:475-480; Bordignon, C. et al. (1995) Science 270:470-475), cystic fibrosis (Zabner, J. et al. (1993) Cell 75:207-216; Crystal, R.G. et al. (1995) Hum. Gene Therapy 6:643-666; Crystal, R.G. et al. (1995) Hum. Gene Therapy 6:667-703), thalassamias, familial hypercholesterolemia, and hemophilia resulting from Factor VIII or Factor IX deficiencies (Crystal, R.G. (1995) Science 270:404-410; Verma, I.M. and N. Somia (1997) Nature 389:239-242)), (ii) express a conditionally lethal gene product (e.g., in the case of cancers which result from unregulated cell proliferation), or (iii) express a protein which affords protection against intracellular parasites (e.g., against human retroviruses, such as human immunodeficiency virus (HIV) (Baltimore, D. (1988) Nature 335:395-396; Poeschla, E. et al. (1996) Proc. Natl. Acad. Sci. USA 93:11395-11399), hepatitis B or C virus (HBV, HCV); fungal parasites, such as *Candida albicans* and *Paracoccidioides brasiliensis*; and protozoan parasites such as *Plasmodium falciparum* and *Trypanosoma cruzi*). In the case where a genetic deficiency in NTRAN expression or regulation causes disease, the expression of NTRAN from an appropriate population of transduced cells may alleviate the clinical manifestations caused by the genetic deficiency.

In a further embodiment of the invention, diseases or disorders caused by deficiencies in NTRAN are treated by constructing mammalian expression vectors encoding NTRAN and introducing these vectors by mechanical means into NTRAN-deficient cells. Mechanical transfer technologies for use with cells *in vivo* or *ex vitro* include (i) direct DNA microinjection into individual cells, (ii) ballistic gold particle delivery, (iii) liposome-mediated transfection, (iv) receptor-mediated gene transfer, and (v) the use of DNA transposons (Morgan, R.A. and W.F. Anderson (1993) Annu. Rev. Biochem. 62:191-217; Ivics, Z. (1997) Cell 91:501-510; Boulay, J.-L. and H. Récipon (1998) Curr. Opin. Biotechnol. 9:445-450).

Expression vectors that may be effective for the expression of NTRAN include, but are not limited to, the PCDNA 3.1, EPITAG, PRCCMV2, PREP, PVAX, PCR2-TOPOTA vectors (Invitrogen, Carlsbad CA), PCMV-SCRIPT, PCMV-TAG, PEGSH/PERV (Stratagene, La Jolla CA), and PTET-OFF, PTET-ON, PTRE2, PTRE2-LUC, PTK-HYG (Clontech, Palo Alto CA). NTRAN may be expressed using (i) a constitutively active promoter, (e.g., from cytomegalovirus (CMV), Rous sarcoma virus (RSV), SV40 virus, thymidine kinase (TK), or β-actin genes), (ii) an inducible promoter (e.g., the tetracycline-regulated promoter (Gossen, M. and H. Bujard (1992) Proc. Natl. Acad. Sci. USA 89:5547-5551; Gossen, M. et al. (1995) Science 268:1766-1769; Rossi, F.M.V. and H.M. Blau (1998) Curr. Opin. Biotechnol. 9:451-456), commercially available in the T-REX plasmid (Invitrogen));

the ecdysone-inducible promoter (available in the plasmids PVGRXR and PIND; Invitrogen); the FK506/rapamycin inducible promoter; or the RU486/mifepristone inducible promoter (Rossi, F.M.V. and H.M. Blau, *supra*)), or (iii) a tissue-specific promoter or the native promoter of the endogenous gene encoding NTRAN from a normal individual.

Commercially available liposome transformation kits (e.g., the PERFECT LIPID TRANSFECTION KIT, available from Invitrogen) allow one with ordinary skill in the art to deliver polynucleotides to target cells in culture and require minimal effort to optimize experimental parameters. In the alternative, transformation is performed using the calcium phosphate method (Graham, F.L. and A.J. Eb (1973) Virology 52:456-467), or by electroporation (Neumann, E. et al. (1982) EMBO J. 1:841-845). The introduction of DNA to primary cells requires modification of these standardized mammalian transfection protocols.

In another embodiment of the invention, diseases or disorders caused by genetic defects with respect to NTRAN expression are treated by constructing a retrovirus vector consisting of (i) the polynucleotide encoding NTRAN under the control of an independent promoter or the retrovirus long terminal repeat (LTR) promoter, (ii) appropriate RNA packaging signals, and (iii) a Rev-responsive element (RRE) along with additional retrovirus *cis*-acting RNA sequences and coding sequences required for efficient vector propagation. Retrovirus vectors (e.g., PFB and PFBNEO) are commercially available (Stratagene) and are based on published data (Riviere, I. et al. (1995) Proc. Natl. Acad. Sci. USA 92:6733-6737), incorporated by reference herein. The vector is propagated in an appropriate vector producing cell line (VPCL) that expresses an envelope gene with a tropism for receptors on the target cells or a promiscuous envelope protein such as VSVg (Armentano, D. et al. (1987) J. Virol. 61:1647-1650; Bender, M.A. et al. (1987) J. Virol. 61:1639-1646; Adam, M.A. and A.D. Miller (1988) J. Virol. 62:3802-3806; Dull, T. et al. (1998) J. Virol. 72:8463-8471; Zufferey, R. et al. (1998) J. Virol. 72:9873-9880). U.S. Patent No. 5,910,434 to Rigg ("Method for obtaining retrovirus packaging cell lines producing high transducing efficiency retroviral supernatant") discloses a method for obtaining retrovirus packaging cell lines and is hereby incorporated by reference. Propagation of retrovirus vectors, transduction of a population of cells (e.g., $CD4^+$ T-cells), and the return of transduced cells to a patient are procedures well known to persons skilled in the art of gene therapy and have been well documented (Ranga, U. et al. (1997) J. Virol. 71:7020-7029; Bauer, G. et al. (1997) Blood 89:2259-2267; Bonyhadi, M.L. (1997) J. Virol. 71:4707-4716; Ranga, U. et al. (1998) Proc. Natl. Acad. Sci. USA 95:1201-1206; Su, L. (1997) Blood 89:2283-2290).

In an embodiment, an adenovirus-based gene therapy delivery system is used to deliver

polynucleotides encoding NTRAN to cells which have one or more genetic abnormalities with respect to the expression of NTRAN. The construction and packaging of adenovirus-based vectors are well known to those with ordinary skill in the art. Replication defective adenovirus vectors have proven to be versatile for importing genes encoding immunoregulatory proteins into intact islets in the pancreas (Csete, M.E. et al. (1995) Transplantation 27:263-268). Potentially useful adenoviral vectors are described in U.S. Patent No. 5,707,618 to Armentano ("Adenovirus vectors for gene therapy"), hereby incorporated by reference. For adenoviral vectors, see also Antinozzi, P.A. et al. (1999; Annu. Rev. Nutr. 19:511-544) and Verma, I.M. and N. Somia (1997; Nature 18:389:239-242).

In another embodiment, a herpes-based, gene therapy delivery system is used to deliver polynucleotides encoding NTRAN to target cells which have one or more genetic abnormalities with respect to the expression of NTRAN. The use of herpes simplex virus (HSV)-based vectors may be especially valuable for introducing NTRAN to cells of the central nervous system, for which HSV has a tropism. The construction and packaging of herpes-based vectors are well known to those with ordinary skill in the art. A replication-competent herpes simplex virus (HSV) type 1-based vector has been used to deliver a reporter gene to the eyes of primates (Liu, X. et al. (1999) Exp. Eye Res. 169:385-395). The construction of a HSV-1 virus vector has also been disclosed in detail in U.S. Patent No. 5,804,413 to DeLuca ("Herpes simplex virus strains for gene transfer"), which is hereby incorporated by reference. U.S. Patent No. 5,804,413 teaches the use of recombinant HSV d92 which consists of a genome containing at least one exogenous gene to be transferred to a cell under the control of the appropriate promoter for purposes including human gene therapy. Also taught by this patent are the construction and use of recombinant HSV strains deleted for ICP4, ICP27 and ICP22. For HSV vectors, see also Goins, W.F. et al. (1999; J. Virol. 73:519-532) and Xu, H. et al. (1994; Dev. Biol. 163:152-161). The manipulation of cloned herpesvirus sequences, the generation of recombinant virus following the transfection of multiple plasmids containing different segments of the large herpesvirus genomes, the growth and propagation of herpesvirus, and the infection of cells with herpesvirus are techniques well known to those of ordinary skill in the art.

In another embodiment, an alphavirus (positive, single-stranded RNA virus) vector is used to deliver polynucleotides encoding NTRAN to target cells. The biology of the prototypic alphavirus, Semliki Forest Virus (SFV), has been studied extensively and gene transfer vectors have been based on the SFV genome (Garoff, H. and K.-J. Li (1998) Curr. Opin. Biotechnol. 9:464-469). During alphavirus RNA replication, a subgenomic RNA is generated that normally encodes the viral capsid proteins. This subgenomic RNA replicates to higher levels than the full length genomic RNA,

resulting in the overproduction of capsid proteins relative to the viral proteins with enzymatic activity (e.g., protease and polymerase). Similarly, inserting the coding sequence for NTRAN into the alphavirus genome in place of the capsid-coding region results in the production of a large number of NTRAN-coding RNAs and the synthesis of high levels of NTRAN in vector transduced cells. While alphavirus infection is typically associated with cell lysis within a few days, the ability to establish a persistent infection in hamster normal kidney cells (BHK-21) with a variant of Sindbis virus (SIN) indicates that the lytic replication of alphaviruses can be altered to suit the needs of the gene therapy application (Dryga, S.A. et al. (1997) Virology 228:74-83). The wide host range of alphaviruses will allow the introduction of NTRAN into a variety of cell types. The specific transduction of a subset of cells in a population may require the sorting of cells prior to transduction. The methods of manipulating infectious cDNA clones of alphaviruses, performing alphavirus cDNA and RNA transfections, and performing alphavirus infections, are well known to those with ordinary skill in the art.

Oligonucleotides derived from the transcription initiation site, e.g., between about positions -10 and +10 from the start site, may also be employed to inhibit gene expression. Similarly, inhibition can be achieved using triple helix base-pairing methodology. Triple helix pairing is useful because it causes inhibition of the ability of the double helix to open sufficiently for the binding of polymerases, transcription factors, or regulatory molecules. Recent therapeutic advances using triplex DNA have been described in the literature (Gee, J.E. et al. (1994) in Huber, B.E. and B.I. Carr, Molecular and Immunologic Approaches, Futura Publishing, Mt. Kisco NY, pp. 163-177). A complementary sequence or antisense molecule may also be designed to block translation of mRNA by preventing the transcript from binding to ribosomes.

Ribozymes, enzymatic RNA molecules, may also be used to catalyze the specific cleavage of RNA. The mechanism of ribozyme action involves sequence-specific hybridization of the ribozyme molecule to complementary target RNA, followed by endonucleolytic cleavage. For example, engineered hammerhead motif ribozyme molecules may specifically and efficiently catalyze endonucleolytic cleavage of RNA molecules encoding NTRAN.

Specific ribozyme cleavage sites within any potential RNA target are initially identified by scanning the target molecule for ribozyme cleavage sites, including the following sequences: GUA, GUU, and GUC. Once identified, short RNA sequences of between 15 and 20 ribonucleotides, corresponding to the region of the target gene containing the cleavage site, may be evaluated for secondary structural features which may render the oligonucleotide inoperable. The suitability of

candidate targets may also be evaluated by testing accessibility to hybridization with complementary oligonucleotides using ribonuclease protection assays.

Complementary ribonucleic acid molecules and ribozymes may be prepared by any method known in the art for the synthesis of nucleic acid molecules. These include techniques for chemically synthesizing oligonucleotides such as solid phase phosphoramidite chemical synthesis. Alternatively, RNA molecules may be generated by *in vitro* and *in vivo* transcription of DNA molecules encoding NTRAN. Such DNA sequences may be incorporated into a wide variety of vectors with suitable RNA polymerase promoters such as T7 or SP6. Alternatively, these cDNA constructs that synthesize complementary RNA, constitutively or inducibly, can be introduced into cell lines, cells, or tissues.

RNA molecules may be modified to increase intracellular stability and half-life. Possible modifications include, but are not limited to, the addition of flanking sequences at the 5' and/or 3' ends of the molecule, or the use of phosphorothioate or 2' O-methyl rather than phosphodiesterase linkages within the backbone of the molecule. This concept is inherent in the production of PNAs and can be extended in all of these molecules by the inclusion of nontraditional bases such as inosine, queosine, and wybutosine, as well as acetyl-, methyl-, thio-, and similarly modified forms of adenine, cytidine, guanine, thymine, and uridine which are not as easily recognized by endogenous endonucleases.

In other embodiments of the invention, the expression of one or more selected polynucleotides of the present invention can be altered, inhibited, decreased, or silenced using RNA interference (RNAi) or post-transcriptional gene silencing (PTGS) methods known in the art. RNAi is a post-transcriptional mode of gene silencing in which double-stranded RNA (dsRNA) introduced into a targeted cell specifically suppresses the expression of the homologous gene (i.e., the gene bearing the sequence complementary to the dsRNA). This effectively knocks out or substantially reduces the expression of the targeted gene. PTGS can also be accomplished by use of DNA or DNA fragments as well. RNAi methods are described by Fire, A. et al. (1998; Nature 391:806-811) and Gura, T. (2000; Nature 404:804-808). PTGS can also be initiated by introduction of a complementary segment of DNA into the selected tissue using gene delivery and/or viral vector delivery methods described herein or known in the art.

RNAi can be induced in mammalian cells by the use of small interfering RNA also known as siRNA. SiRNA are shorter segments of dsRNA (typically about 21 to 23 nucleotides in length) that result *in vivo* from cleavage of introduced dsRNA by the action of an endogenous ribonuclease. SiRNA appear to be the mediators of the RNAi effect in mammals. The most effective siRNAs

appear to be 21 nucleotide dsRNAs with 2 nucleotide 3' overhangs. The use of siRNA for inducing RNAi in mammalian cells is described by Elbashir, S.M. et al. (2001; Nature 411:494-498).

SiRNA can either be generated indirectly by introduction of dsRNA into the targeted cell, or directly by mammalian transfection methods and agents described herein or known in the art (such as liposome-mediated transfection, viral vector methods, or other polynucleotide delivery/introductory methods). Suitable SiRNAs can be selected by examining a transcript of the target polynucleotide (e.g., mRNA) for nucleotide sequences downstream from the AUG start codon and recording the occurrence of each nucleotide and the 3' adjacent 19 to 23 nucleotides as potential siRNA target sites, with sequences having a 21 nucleotide length being preferred. Regions to be avoided for target siRNA sites include the 5' and 3' untranslated regions (UTRs) and regions near the start codon (within 75 bases), as these may be richer in regulatory protein binding sites. UTR-binding proteins and/or translation initiation complexes may interfere with binding of the siRNP endonuclease complex. The selected target sites for siRNA can then be compared to the appropriate genome database (e.g., human, etc.) using BLAST or other sequence comparison algorithms known in the art. Target sequences with significant homology to other coding sequences can be eliminated from consideration. The selected SiRNAs can be produced by chemical synthesis methods known in the art or by *in vitro* transcription using commercially available methods and kits such as the SILENCER siRNA construction kit (Ambion, Austin TX).

In alternative embodiments, long-term gene silencing and/or RNAi effects can be induced in selected tissue using expression vectors that continuously express siRNA. This can be accomplished using expression vectors that are engineered to express hairpin RNAs (shRNAs) using methods known in the art (see, e.g., Brummelkamp, T.R. et al. (2002) Science 296:550-553; and Paddison, P.J. et al. (2002) Genes Dev. 16:948-958). In these and related embodiments, shRNAs can be delivered to target cells using expression vectors known in the art. An example of a suitable expression vector for delivery of siRNA is the PSILENCER1.0-U6 (circular) plasmid (Ambion). Once delivered to the target tissue, shRNAs are processed *in vivo* into siRNA-like molecules capable of carrying out gene-specific silencing.

In various embodiments, the expression levels of genes targeted by RNAi or PTGS methods can be determined by assays for mRNA and/or protein analysis. Expression levels of the mRNA of a targeted gene, can be determined by northern analysis methods using, for example, the NORTHERNMAX-GLY kit (Ambion); by microarray methods; by PCR methods; by real time PCR methods; and by other RNA/polynucleotide assays known in the art or described herein. Expression

levels of the protein encoded by the targeted gene can be determined by Western analysis using standard techniques known in the art.

An additional embodiment of the invention encompasses a method for screening for a compound which is effective in altering expression of a polynucleotide encoding NTRAN. Compounds which may be effective in altering expression of a specific polynucleotide may include, but are not limited to, oligonucleotides, antisense oligonucleotides, triple helix-forming oligonucleotides, transcription factors and other polypeptide transcriptional regulators, and non-macromolecular chemical entities which are capable of interacting with specific polynucleotide sequences. Effective compounds may alter polynucleotide expression by acting as either inhibitors or promoters of polynucleotide expression. Thus, in the treatment of disorders associated with increased NTRAN expression or activity, a compound which specifically inhibits expression of the polynucleotide encoding NTRAN may be therapeutically useful, and in the treatment of disorders associated with decreased NTRAN expression or activity, a compound which specifically promotes expression of the polynucleotide encoding NTRAN may be therapeutically useful.

In various embodiments, one or more test compounds may be screened for effectiveness in altering expression of a specific polynucleotide. A test compound may be obtained by any method commonly known in the art, including chemical modification of a compound known to be effective in altering polynucleotide expression; selection from an existing, commercially-available or proprietary library of naturally-occurring or non-natural chemical compounds; rational design of a compound based on chemical and/or structural properties of the target polynucleotide; and selection from a library of chemical compounds created combinatorially or randomly. A sample comprising a polynucleotide encoding NTRAN is exposed to at least one test compound thus obtained. The sample may comprise, for example, an intact or permeabilized cell, or an *in vitro* cell-free or reconstituted biochemical system. Alterations in the expression of a polynucleotide encoding NTRAN are assayed by any method commonly known in the art. Typically, the expression of a specific nucleotide is detected by hybridization with a probe having a nucleotide sequence complementary to the sequence of the polynucleotide encoding NTRAN. The amount of hybridization may be quantified, thus forming the basis for a comparison of the expression of the polynucleotide both with and without exposure to one or more test compounds. Detection of a change in the expression of a polynucleotide exposed to a test compound indicates that the test compound is effective in altering the expression of the polynucleotide. A screen for a compound effective in altering expression of a specific polynucleotide can be carried out, for example, using a *Schizosaccharomyces pombe* gene expression system

(Atkins, D. et al. (1999) U.S. Patent No. 5,932,435; Arndt, G.M. et al. (2000) Nucleic Acids Res. 28:E15) or a human cell line such as HeLa cell (Clarke, M.L. et al. (2000) Biochem. Biophys. Res. Commun. 268:8-13). A particular embodiment of the present invention involves screening a combinatorial library of oligonucleotides (such as deoxyribonucleotides, ribonucleotides, peptide nucleic acids, and modified oligonucleotides) for antisense activity against a specific polynucleotide sequence (Bruice, T.W. et al. (1997) U.S. Patent No. 5,686,242; Bruice, T.W. et al. (2000) U.S. Patent No. 6,022,691).

Many methods for introducing vectors into cells or tissues are available and equally suitable for use *in vivo*, *in vitro*, and *ex vivo*. For *ex vivo* therapy, vectors may be introduced into stem cells taken from the patient and clonally propagated for autologous transplant back into that same patient. Delivery by transfection, by liposome injections, or by polycationic amino polymers may be achieved using methods which are well known in the art (Goldman, C.K. et al. (1997) Nat. Biotechnol. 15:462-466).

Any of the therapeutic methods described above may be applied to any subject in need of such therapy, including, for example, mammals such as humans, dogs, cats, cows, horses, rabbits, and monkeys.

An additional embodiment of the invention relates to the administration of a composition which generally comprises an active ingredient formulated with a pharmaceutically acceptable excipient. Excipients may include, for example, sugars, starches, celluloses, gums, and proteins. Various formulations are commonly known and are thoroughly discussed in the latest edition of Remington's Pharmaceutical Sciences (Maack Publishing, Easton PA). Such compositions may consist of NTRAN, antibodies to NTRAN, and mimetics, agonists, antagonists, or inhibitors of NTRAN.

In various embodiments, the compositions described herein, such as pharmaceutical compositions, may be administered by any number of routes including, but not limited to, oral, intravenous, intramuscular, intra-arterial, intramedullary, intrathecal, intraventricular, pulmonary, transdermal, subcutaneous, intraperitoneal, intranasal, enteral, topical, sublingual, or rectal means.

Compositions for pulmonary administration may be prepared in liquid or dry powder form. These compositions are generally aerosolized immediately prior to inhalation by the patient. In the case of small molecules (e.g. traditional low molecular weight organic drugs), aerosol delivery of fast-acting formulations is well-known in the art. In the case of macromolecules (e.g. larger peptides and proteins), recent developments in the field of pulmonary delivery via the alveolar region of the lung have enabled the practical delivery of drugs such as insulin to blood circulation (see, e.g., Patton, J.S.

et al., U.S. Patent No. 5,997,848). Pulmonary delivery allows administration without needle injection, and obviates the need for potentially toxic penetration enhancers.

Compositions suitable for use in the invention include compositions wherein the active ingredients are contained in an effective amount to achieve the intended purpose. The determination of an effective dose is well within the capability of those skilled in the art.

Specialized forms of compositions may be prepared for direct intracellular delivery of macromolecules comprising NTRAN or fragments thereof. For example, liposome preparations containing a cell-impermeable macromolecule may promote cell fusion and intracellular delivery of the macromolecule. Alternatively, NTRAN or a fragment thereof may be joined to a short cationic N-terminal portion from the HIV Tat-1 protein. Fusion proteins thus generated have been found to transduce into the cells of all tissues, including the brain, in a mouse model system (Schwarze, S.R. et al. (1999) Science 285:1569-1572).

For any compound, the therapeutically effective dose can be estimated initially either in cell culture assays, e.g., of neoplastic cells, or in animal models such as mice, rats, rabbits, dogs, monkeys, or pigs. An animal model may also be used to determine the appropriate concentration range and route of administration. Such information can then be used to determine useful doses and routes for administration in humans.

A therapeutically effective dose refers to that amount of active ingredient, for example NTRAN or fragments thereof, antibodies of NTRAN, and agonists, antagonists or inhibitors of NTRAN, which ameliorates the symptoms or condition. Therapeutic efficacy and toxicity may be determined by standard pharmaceutical procedures in cell cultures or with experimental animals, such as by calculating the ED_{50} (the dose therapeutically effective in 50% of the population) or LD_{50} (the dose lethal to 50% of the population) statistics. The dose ratio of toxic to therapeutic effects is the therapeutic index, which can be expressed as the LD_{50}/ED_{50} ratio. Compositions which exhibit large therapeutic indices are preferred. The data obtained from cell culture assays and animal studies are used to formulate a range of dosage for human use. The dosage contained in such compositions is preferably within a range of circulating concentrations that includes the ED_{50} with little or no toxicity. The dosage varies within this range depending upon the dosage form employed, the sensitivity of the patient, and the route of administration.

The exact dosage will be determined by the practitioner, in light of factors related to the subject requiring treatment. Dosage and administration are adjusted to provide sufficient levels of the active moiety or to maintain the desired effect. Factors which may be taken into account include the

severity of the disease state, the general health of the subject, the age, weight, and gender of the subject, time and frequency of administration, drug combination(s), reaction sensitivities, and response to therapy. Long-acting compositions may be administered every 3 to 4 days, every week, or biweekly depending on the half-life and clearance rate of the particular formulation.

Normal dosage amounts may vary from about 0.1 μg to 100,000 μg, up to a total dose of about 1 gram, depending upon the route of administration. Guidance as to particular dosages and methods of delivery is provided in the literature and generally available to practitioners in the art. Those skilled in the art will employ different formulations for nucleotides than for proteins or their inhibitors. Similarly, delivery of polynucleotides or polypeptides will be specific to particular cells, conditions, locations, etc.

DIAGNOSTICS

In another embodiment, antibodies which specifically bind NTRAN may be used for the diagnosis of disorders characterized by expression of NTRAN, or in assays to monitor patients being treated with NTRAN or agonists, antagonists, or inhibitors of NTRAN. Antibodies useful for diagnostic purposes may be prepared in the same manner as described above for therapeutics. Diagnostic assays for NTRAN include methods which utilize the antibody and a label to detect NTRAN in human body fluids or in extracts of cells or tissues. The antibodies may be used with or without modification, and may be labeled by covalent or non-covalent attachment of a reporter molecule. A wide variety of reporter molecules, several of which are described above, are known in the art and may be used.

A variety of protocols for measuring NTRAN, including ELISAs, RIAs, and FACS, are known in the art and provide a basis for diagnosing altered or abnormal levels of NTRAN expression. Normal or standard values for NTRAN expression are established by combining body fluids or cell extracts taken from normal mammalian subjects, for example, human subjects, with antibodies to NTRAN under conditions suitable for complex formation. The amount of standard complex formation may be quantitated by various methods, such as photometric means. Quantities of NTRAN expressed in subject, control, and disease samples from biopsied tissues are compared with the standard values. Deviation between standard and subject values establishes the parameters for diagnosing disease.

In another embodiment of the invention, polynucleotides encoding NTRAN may be used for diagnostic purposes. The polynucleotides which may be used include oligonucleotides, complementary RNA and DNA molecules, and PNAs. The polynucleotides may be used to detect and quantify gene expression in biopsied tissues in which expression of NTRAN may be correlated with disease. The

diagnostic assay may be used to determine absence, presence, and excess expression of NTRAN, and to monitor regulation of NTRAN levels during therapeutic intervention.

In one aspect, hybridization with PCR probes which are capable of detecting polynucleotides, including genomic sequences, encoding NTRAN or closely related molecules may be used to identify nucleic acid sequences which encode NTRAN. The specificity of the probe, whether it is made from a highly specific region, e.g., the 5' regulatory region, or from a less specific region, e.g., a conserved motif, and the stringency of the hybridization or amplification will determine whether the probe identifies only naturally occurring sequences encoding NTRAN, allelic variants, or related sequences.

Probes may also be used for the detection of related sequences, and may have at least 50% sequence identity to any of the NTRAN encoding sequences. The hybridization probes of the subject invention may be DNA or RNA and may be derived from the sequence of SEQ ID NO:23-44 or from genomic sequences including promoters, enhancers, and introns of the NTRAN gene.

Means for producing specific hybridization probes for polynucleotides encoding NTRAN include the cloning of polynucleotides encoding NTRAN or NTRAN derivatives into vectors for the production of mRNA probes. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes *in vitro* by means of the addition of the appropriate RNA polymerases and the appropriate labeled nucleotides. Hybridization probes may be labeled by a variety of reporter groups, for example, by radionuclides such as ^{32}P or ^{35}S, or by enzymatic labels, such as alkaline phosphatase coupled to the probe via avidin/biotin coupling systems, and the like.

Polynucleotides encoding NTRAN may be used for the diagnosis of disorders associated with expression of NTRAN. Examples of such disorders include, but are not limited to, an autoimmune/inflammatory disorder such as acquired immunodeficiency syndrome (AIDS), Addison's disease, adult respiratory distress syndrome, allergies, ankylosing spondylitis, amyloidosis, anemia, asthma, atherosclerosis, autoimmune hemolytic anemia, autoimmune thyroiditis, autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED), bronchitis, cholecystitis, contact dermatitis, Crohn's disease, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, episodic lymphopenia with lymphocytotoxins, erythroblastosis fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis, hypereosinophilia, irritable bowel syndrome, multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis, thromb osteoarthritis, osteoporosis, pancreatitis, polymyositis, psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjögren's ocytopenic purpura, ulcerative colitis, uveitis, Werner

syndrome, complications of cancer, hemodialysis, and extracorporeal circulation, viral, bacterial, fungal, parasitic, protozoal, and helminthic infections, and trauma; a cardiovascular disorder such as congestive heart failure, ischemic heart disease, angina pectoris, myocardial infarction, hypertensive heart disease, degenerative valvular heart disease, calcific aortic valve stenosis, congenitally bicuspid aortic valve, mitral annular calcification, mitral valve prolapse, rheumatic fever and rheumatic heart disease, infective endocarditis, nonbacterial thrombotic endocarditis, endocarditis of systemic lupus erythematosus, carcinoid heart disease, cardiomyopathy, myocarditis, pericarditis, neoplastic heart disease, congenital heart disease, and complications of cardiac transplantation, arteriovenous fistula, atherosclerosis, hypertension, vasculitis, Raynaud's disease, aneurysms, arterial dissections, varicose veins, thrombophlebitis and phlebothrombosis, vascular tumors, and complications of thrombolysis, balloon angioplasty, vascular replacement, and coronary artery bypass graft surgery; a neurological disorder such as epilepsy, ischemic cerebrovascular disease, stroke, cerebral neoplasms, Pick's disease, Huntington's disease, dementia, Parkinson's disease and other extrapyramidal disorders, amyotrophic lateral sclerosis and other motor neuron disorders, progressive neural muscular atrophy, retinitis pigmentosa, hereditary ataxias, multiple sclerosis and other demyelinating diseases, bacterial and viral meningitis, brain abscess, subdural empyema, epidural abscess, suppurative intracranial thrombophlebitis, myelitis and radiculitis, viral central nervous system disease, prion diseases including kuru, Creutzfeldt-Jakob disease, and Gerstmann-Straussler-Scheinker syndrome, fatal familial insomnia, nutritional and metabolic diseases of the nervous system, neurofibromatosis, tuberous sclerosis, cerebelloretinal hemangioblastomatosis, encephalotrigeminal syndrome, mental retardation and other developmental disorders of the central nervous system including Down syndrome, cerebral palsy, neuroskeletal disorders, autonomic nervous system disorders, cranial nerve disorders, spinal cord diseases, muscular dystrophy and other neuromuscular disorders, peripheral nervous system disorders, dermatomyositis and polymyositis, inherited, metabolic, endocrine, and toxic myopathies, myasthenia gravis, periodic paralysis, mental disorders including mood, anxiety, and schizophrenic disorders, seasonal affective disorder (SAD), akathesia, amnesia, catatonia, diabetic neuropathy, tardive dyskinesia, dystonias, paranoid psychoses, postherpetic neuralgia, Tourette's disorder, progressive supranuclear palsy, corticobasal degeneration, and familial frontotemporal dementia; a developmental disorder such as renal tubular acidosis, anemia, Cushing's syndrome, achondroplastic dwarfism, Duchenne and Becker muscular dystrophy, epilepsy, gonadal dysgenesis, WAGR syndrome (Wilms' tumor, aniridia, genitourinary abnormalities, and mental retardation), Smith-Magenis syndrome, myelodysplastic syndrome, hereditary mucoepithelial dysplasia, hereditary keratodermas, hereditary

neuropathies such as Charcot-Marie-Tooth disease and neurofibromatosis, hypothyroidism, hydrocephalus, seizure disorders such as Syndenham's chorea and cerebral palsy, spina bifida, anencephaly, craniorachischisis, congenital glaucoma, cataract, and sensorineural hearing loss; a cell proliferative disorder such as actinic keratosis, arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia, and cancers including adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, cancers of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, colon, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus and a cancer such as adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, cancers of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus; a transport disorder such as akinesia, amyotrophic lateral sclerosis, ataxia telangiectasia, cystic fibrosis, Becker's muscular dystrophy, Bell's palsy, Charcot-Marie Tooth disease, diabetes mellitus, diabetes insipidus, diabetic neuropathy, Duchenne muscular dystrophy, hyperkalemic periodic paralysis, normokalemic periodic paralysis, Parkinson's disease, malignant hyperthermia, multidrug resistance, myasthenia gravis, myotonic dystrophy, catatonia, tardive dyskinesia, dystonias, peripheral neuropathy, cerebral neoplasms, prostate cancer, cardiac disorders associated with transport, e.g., angina, bradyarrythmia, tachyarrythmia, hypertension, Long QT syndrome, myocarditis, cardiomyopathy, nemaline myopathy, centronuclear myopathy, lipid myopathy, mitochondrial myopathy, thyrotoxic myopathy, ethanol myopathy, dermatomyositis, inclusion body myositis, infectious myositis, polymyositis, neurological disorders associated with transport, e.g., Alzheimer's disease, amnesia, bipolar disorder, dementia, depression, epilepsy, Tourette's disorder, paranoid psychoses, and schizophrenia, and other disorders associated with transport, e.g., neurofibromatosis, postherpetic neuralgia, trigeminal neuropathy, sarcoidosis, sickle cell anemia, Wilson's disease, cataracts, infertility, pulmonary artery stenosis, sensorineural autosomal deafness, hyperglycemia, hypoglycemia, Grave's disease, goiter, Cushing's disease, Addison's disease, glucose-galactose malabsorption syndrome, hypercholesterolemia, adrenoleukodystrophy, Zellweger syndrome, Menkes disease, occipital horn syndrome, von Gierke disease, cystinuria, iminoglycinuria, Hartup disease, and Fanconi disease; a psychiatric disorder such as acute stress disorder, alcohol dependence, amphetamine dependence, anorexia nervosa, antisocial

personality disorder, attention-deficit hyperactivity disorder, autistic disorder, anxiety, avoidant personality disorder, bipolar disorder, borderline personality disorder, brief psychotic disorder, bulimia nervosa, cannabis dependence, cocaine dependence, conduct disorder, cyclothymic disorder, delirium, delusional disorder, dementia, dependent personality disorder, depression, dysthymic disorder, hallucinogen dependence, histrionic personality disorder, inhalant dependence, manic depression, multi-infarct dementia, narcissistic personality disorder, nicotine dependence, obsessive-compulsive disorder, opioid dependence, oppositional defiant disorder, panic disorder, paranoid personality disorder, phencyclidine dependence, phobia, posttraumatic stress disorder, schizoaffective disorder, schizoid personality disorder, schizophrenia, sedative dependence, separation anxiety disorder, and sleep disorder; a metabolic disorder such as Addison's disease, cerebrotendinous xanthomatosis, congenital adrenal hyperplasia, coumarin resistance, cystic fibrosis, fatty hepatocirrhosis, fructose-1,6-diphosphatase deficiency, galactosemia, goiter, glucagonoma, glycogen storage diseases, hereditary fructose intolerance, hyperadrenalism, hypoadrenalism, hyperparathyroidism, hypoparathyroidism, hypercholesterolemia, hyperthyroidism, hypoglycemia, hypothyroidism, hyperlipidemia, hyperlipemia, lipid myopathies, lipodystrophies, lysosomal storage diseases, mannosidosis, neuraminidase deficiency, obesity, osteoporosis, phenylketonuria, pseudovitamin D-deficiency rickets, disorders of carbohydrate metabolism such as congenital type II dyserythropoietic anemia, diabetes, insulin-dependent diabetes mellitus, non-insulin-dependent diabetes mellitus, galactose epimerase deficiency, glycogen storage diseases, lysosomal storage diseases, fructosuria, pentosuria, and inherited abnormalities of pyruvate metabolism, disorders of lipid metabolism such as fatty liver, cholestasis, primary biliary cirrhosis, carnitine deficiency, carnitine palmitoyltransferase deficiency, myoadenylate deaminase deficiency, hypertriglyceridemia, lipid storage disorders such Fabry's disease, Gaucher's disease, Niemann-Pick's disease, metachromatic leukodystrophy, adrenoleukodystrophy, GM_2 gangliosidosis, and ceroid lipofuscinosis, abetalipoproteinemia, Tangier disease, hyperlipoproteinemia, lipodystrophy, lipomatoses, acute panniculitis, disseminated fat necrosis, adiposis dolorosa, lipoid adrenal hyperplasia, minimal change disease, lipomas, atherosclerosis, hypercholesterolemia, hypercholesterolemia with hypertriglyceridemia, primary hypoalphalipoproteinemia, hypothyroidism, renal disease, liver disease, lecithin:cholesterol acyltransferase deficiency, cerebrotendinous xanthomatosis, sitosterolemia, hypocholesterolemia, Tay-Sachs disease, Sandhoff's disease, hyperlipidemia, hyperlipemia, and lipid myopathies, and disorders of copper metabolism such as Menke's disease, Wilson's disease, and Ehlers-Danlos syndrome type IX diabetes; and an endocrine disorder such as a disorder of the hypothalamus and/or pituitary resulting

from lesions such as a primary brain tumor, adenoma, infarction associated with pregnancy, hypophysectomy, aneurysm, vascular malformation, thrombosis, infection, immunological disorder, and complication due to head trauma, a disorder associated with hypopituitarism including hypogonadism, Sheehan syndrome, diabetes insipidus, Kallman's disease, Hand-Schuller-Christian disease, Letterer-Siwe disease, sarcoidosis, empty sella syndrome, and dwarfism, a disorder associated with hyperpituitarism including acromegaly, giantism, and syndrome of inappropriate antidiuretic hormone (ADH) secretion (SIADH) often caused by benign adenoma, a disorder associated with hypothyroidism including goiter, myxedema, acute thyroiditis associated with bacterial infection, subacute thyroiditis associated with viral infection, autoimmune thyroiditis (Hashimoto's disease), and cretinism, a disorder associated with hyperthyroidism including thyrotoxicosis and its various forms, Grave's disease, pretibial myxedema, toxic multinodular goiter, thyroid carcinoma, and Plummer's disease, a disorder associated with hyperparathyroidism including Conn disease (chronic hypercalemia), a pancreatic disorder such as Type I or Type II diabetes mellitus and associated complications, a disorder associated with the adrenals such as hyperplasia, carcinoma, or adenoma of the adrenal cortex, hypertension associated with alkalosis, amyloidosis, hypokalemia, Cushing's disease, Liddle's syndrome, and Arnold-Healy-Gordon syndrome, pheochromocytoma tumors, and Addison's disease, a disorder associated with gonadal steroid hormones such as: in women, abnormal prolactin production, infertility, endometriosis, perturbation of the menstrual cycle, polycystic ovarian disease, hyperprolactinemia, isolated gonadotropin deficiency, amenorrhea, galactorrhea, hermaphroditism, hirsutism and virilization, breast cancer, and, in post-menopausal women, osteoporosis, and, in men, Leydig cell deficiency, male climacteric phase, and germinal cell aplasia, a hypergonadal disorder associated with Leydig cell tumors, androgen resistance associated with absence of androgen receptors, syndrome of 5 α-reductase, and gynecomastia. Polynucleotides encoding NTRAN may be used in Southern or northern analysis, dot blot, or other membrane-based technologies; in PCR technologies; in dipstick, pin, and multiformat ELISA-like assays; and in microarrays utilizing fluids or tissues from patients to detect altered NTRAN expression. Such qualitative or quantitative methods are well known in the art.

In a particular embodiment, polynucleotides encoding NTRAN may be used in assays that detect the presence of associated disorders, particularly those mentioned above. Polynucleotides complementary to sequences encoding NTRAN may be labeled by standard methods and added to a fluid or tissue sample from a patient under conditions suitable for the formation of hybridization complexes. After a suitable incubation period, the sample is washed and the signal is quantified and

compared with a standard value. If the amount of signal in the patient sample is significantly altered in comparison to a control sample then the presence of altered levels of polynucleotides encoding NTRAN in the sample indicates the presence of the associated disorder. Such assays may also be used to evaluate the efficacy of a particular therapeutic treatment regimen in animal studies, in clinical trials, or to monitor the treatment of an individual patient.

In order to provide a basis for the diagnosis of a disorder associated with expression of NTRAN, a normal or standard profile for expression is established. This may be accomplished by combining body fluids or cell extracts taken from normal subjects, either animal or human, with a sequence, or a fragment thereof, encoding NTRAN, under conditions suitable for hybridization or amplification. Standard hybridization may be quantified by comparing the values obtained from normal subjects with values from an experiment in which a known amount of a substantially purified polynucleotide is used. Standard values obtained in this manner may be compared with values obtained from samples from patients who are symptomatic for a disorder. Deviation from standard values is used to establish the presence of a disorder.

Once the presence of a disorder is established and a treatment protocol is initiated, hybridization assays may be repeated on a regular basis to determine if the level of expression in the patient begins to approximate that which is observed in the normal subject. The results obtained from successive assays may be used to show the efficacy of treatment over a period ranging from several days to months.

With respect to cancer, the presence of an abnormal amount of transcript (either under- or overexpressed) in biopsied tissue from an individual may indicate a predisposition for the development of the disease, or may provide a means for detecting the disease prior to the appearance of actual clinical symptoms. A more definitive diagnosis of this type may allow health professionals to employ preventative measures or aggressive treatment earlier, thereby preventing the development or further progression of the cancer.

Additional diagnostic uses for oligonucleotides designed from the sequences encoding NTRAN may involve the use of PCR. These oligomers may be chemically synthesized, generated enzymatically, or produced *in vitro*. Oligomers will preferably contain a fragment of a polynucleotide encoding NTRAN, or a fragment of a polynucleotide complementary to the polynucleotide encoding NTRAN, and will be employed under optimized conditions for identification of a specific gene or condition. Oligomers may also be employed under less stringent conditions for detection or quantification of closely related DNA or RNA sequences.

In a particular aspect, oligonucleotide primers derived from polynucleotides encoding NTRAN may be used to detect single nucleotide polymorphisms (SNPs). SNPs are substitutions, insertions and deletions that are a frequent cause of inherited or acquired genetic disease in humans. Methods of SNP detection include, but are not limited to, single-stranded conformation polymorphism (SSCP) and fluorescent SSCP (fSSCP) methods. In SSCP, oligonucleotide primers derived from polynucleotides encoding NTRAN are used to amplify DNA using the polymerase chain reaction (PCR). The DNA may be derived, for example, from diseased or normal tissue, biopsy samples, bodily fluids, and the like. SNPs in the DNA cause differences in the secondary and tertiary structures of PCR products in single-stranded form, and these differences are detectable using gel electrophoresis in non-denaturing gels. In fSCCP, the oligonucleotide primers are fluorescently labeled, which allows detection of the amplimers in high-throughput equipment such as DNA sequencing machines. Additionally, sequence database analysis methods, termed in silico SNP (isSNP), are capable of identifying polymorphisms by comparing the sequence of individual overlapping DNA fragments which assemble into a common consensus sequence. These computer-based methods filter out sequence variations due to laboratory preparation of DNA and sequencing errors using statistical models and automated analyses of DNA sequence chromatograms. In the alternative, SNPs may be detected and characterized by mass spectrometry using, for example, the high throughput MASSARRAY system (Sequenom, Inc., San Diego CA).

SNPs may be used to study the genetic basis of human disease. For example, at least 16 common SNPs have been associated with non-insulin-dependent diabetes mellitus. SNPs are also useful for examining differences in disease outcomes in monogenic disorders, such as cystic fibrosis, sickle cell anemia, or chronic granulomatous disease. For example, variants in the mannose-binding lectin, MBL2, have been shown to be correlated with deleterious pulmonary outcomes in cystic fibrosis. SNPs also have utility in pharmacogenomics, the identification of genetic variants that influence a patient's response to a drug, such as life-threatening toxicity. For example, a variation in N-acetyl transferase is associated with a high incidence of peripheral neuropathy in response to the anti-tuberculosis drug isoniazid, while a variation in the core promoter of the ALOX5 gene results in diminished clinical response to treatment with an anti-asthma drug that targets the 5-lipoxygenase pathway. Analysis of the distribution of SNPs in different populations is useful for investigating genetic drift, mutation, recombination, and selection, as well as for tracing the origins of populations and their migrations (Taylor, J.G. et al. (2001) Trends Mol. Med. 7:507-512; Kwok, P.-Y. and Z. Gu (1999) Mol. Med. Today 5:538-543; Nowotny, P. et al. (2001) Curr. Opin. Neurobiol. 11:637-641).

Methods which may also be used to quantify the expression of NTRAN include radiolabeling or biotinylating nucleotides, coamplification of a control nucleic acid, and interpolating results from standard curves (Melby, P.C. et al. (1993) J. Immunol. Methods 159:235-244; Duplaa, C. et al. (1993) Anal. Biochem. 212:229-236). The speed of quantitation of multiple samples may be accelerated by running the assay in a high-throughput format where the oligomer or polynucleotide of interest is presented in various dilutions and a spectrophotometric or colorimetric response gives rapid quantitation.

In further embodiments, oligonucleotides or longer fragments derived from any of the polynucleotides described herein may be used as elements on a microarray. The microarray can be used in transcript imaging techniques which monitor the relative expression levels of large numbers of genes simultaneously as described below. The microarray may also be used to identify genetic variants, mutations, and polymorphisms. This information may be used to determine gene function, to understand the genetic basis of a disorder, to diagnose a disorder, to monitor progression/regression of disease as a function of gene expression, and to develop and monitor the activities of therapeutic agents in the treatment of disease. In particular, this information may be used to develop a pharmacogenomic profile of a patient in order to select the most appropriate and effective treatment regimen for that patient. For example, therapeutic agents which are highly effective and display the fewest side effects may be selected for a patient based on his/her pharmacogenomic profile.

In another embodiment, NTRAN, fragments of NTRAN, or antibodies specific for NTRAN may be used as elements on a microarray. The microarray may be used to monitor or measure protein-protein interactions, drug-target interactions, and gene expression profiles, as described above.

A particular embodiment relates to the use of the polynucleotides of the present invention to generate a transcript image of a tissue or cell type. A transcript image represents the global pattern of gene expression by a particular tissue or cell type. Global gene expression patterns are analyzed by quantifying the number of expressed genes and their relative abundance under given conditions and at a given time (Seilhamer et al., "Comparative Gene Transcript Analysis," U.S. Patent No. 5,840,484; hereby expressly incorporated by reference herein). Thus a transcript image may be generated by hybridizing the polynucleotides of the present invention or their complements to the totality of transcripts or reverse transcripts of a particular tissue or cell type. In one embodiment, the hybridization takes place in high-throughput format, wherein the polynucleotides of the present invention or their complements comprise a subset of a plurality of elements on a microarray. The resultant transcript image would provide a profile of gene activity.

Transcript images may be generated using transcripts isolated from tissues, cell lines, biopsies, or other biological samples. The transcript image may thus reflect gene expression *in vivo*, as in the case of a tissue or biopsy sample, or *in vitro*, as in the case of a cell line.

Transcript images which profile the expression of the polynucleotides of the present invention may also be used in conjunction with *in vitro* model systems and preclinical evaluation of pharmaceuticals, as well as toxicological testing of industrial and naturally-occurring environmental compounds. All compounds induce characteristic gene expression patterns, frequently termed molecular fingerprints or toxicant signatures, which are indicative of mechanisms of action and toxicity (Nuwaysir, E.F. et al. (1999) Mol. Carcinog. 24:153-159; Steiner, S. and N.L. Anderson (2000) Toxicol. Lett. 112-113:467-471). If a test compound has a signature similar to that of a compound with known toxicity, it is likely to share those toxic properties. These fingerprints or signatures are most useful and refined when they contain expression information from a large number of genes and gene families. Ideally, a genome-wide measurement of expression provides the highest quality signature. Even genes whose expression is not altered by any tested compounds are important as well, as the levels of expression of these genes are used to normalize the rest of the expression data. The normalization procedure is useful for comparison of expression data after treatment with different compounds. While the assignment of gene function to elements of a toxicant signature aids in interpretation of toxicity mechanisms, knowledge of gene function is not necessary for the statistical matching of signatures which leads to prediction of toxicity (see, for example, Press Release 00-02 from the National Institute of Environmental Health Sciences, released February 29, 2000, available at http://www.niehs.nih.gov/oc/news/toxchip.htm). Therefore, it is important and desirable in toxicological screening using toxicant signatures to include all expressed gene sequences.

In an embodiment, the toxicity of a test compound can be assessed by treating a biological sample containing nucleic acids with the test compound. Nucleic acids that are expressed in the treated biological sample are hybridized with one or more probes specific to the polynucleotides of the present invention, so that transcript levels corresponding to the polynucleotides of the present invention may be quantified. The transcript levels in the treated biological sample are compared with levels in an untreated biological sample. Differences in the transcript levels between the two samples are indicative of a toxic response caused by the test compound in the treated sample.

Another embodiment relates to the use of the polypeptides disclosed herein to analyze the proteome of a tissue or cell type. The term proteome refers to the global pattern of protein expression in a particular tissue or cell type. Each protein component of a proteome can be subjected individually

to further analysis. Proteome expression patterns, or profiles, are analyzed by quantifying the number of expressed proteins and their relative abundance under given conditions and at a given time. A profile of a cell's proteome may thus be generated by separating and analyzing the polypeptides of a particular tissue or cell type. In one embodiment, the separation is achieved using two-dimensional gel electrophoresis, in which proteins from a sample are separated by isoelectric focusing in the first dimension, and then according to molecular weight by sodium dodecyl sulfate slab gel electrophoresis in the second dimension (Steiner and Anderson, *supra*). The proteins are visualized in the gel as discrete and uniquely positioned spots, typically by staining the gel with an agent such as Coomassie Blue or silver or fluorescent stains. The optical density of each protein spot is generally proportional to the level of the protein in the sample. The optical densities of equivalently positioned protein spots from different samples, for example, from biological samples either treated or untreated with a test compound or therapeutic agent, are compared to identify any changes in protein spot density related to the treatment. The proteins in the spots are partially sequenced using, for example, standard methods employing chemical or enzymatic cleavage followed by mass spectrometry. The identity of the protein in a spot may be determined by comparing its partial sequence, preferably of at least 5 contiguous amino acid residues, to the polypeptide sequences of interest. In some cases, further sequence data may be obtained for definitive protein identification.

A proteomic profile may also be generated using antibodies specific for NTRAN to quantify the levels of NTRAN expression. In one embodiment, the antibodies are used as elements on a microarray, and protein expression levels are quantified by exposing the microarray to the sample and detecting the levels of protein bound to each array element (Lueking, A. et al. (1999) Anal. Biochem. 270:103-111; Mendoze, L.G. et al. (1999) Biotechniques 27:778-788). Detection may be performed by a variety of methods known in the art, for example, by reacting the proteins in the sample with a thiol- or amino-reactive fluorescent compound and detecting the amount of fluorescence bound at each array element.

Toxicant signatures at the proteome level are also useful for toxicological screening, and should be analyzed in parallel with toxicant signatures at the transcript level. There is a poor correlation between transcript and protein abundances for some proteins in some tissues (Anderson, N.L. and J. Seilhamer (1997) Electrophoresis 18:533-537), so proteome toxicant signatures may be useful in the analysis of compounds which do not significantly affect the transcript image, but which alter the proteomic profile. In addition, the analysis of transcripts in body fluids is difficult, due to rapid degradation of mRNA, so proteomic profiling may be more reliable and informative in such cases.

In another embodiment, the toxicity of a test compound is assessed by treating a biological sample containing proteins with the test compound. Proteins that are expressed in the treated biological sample are separated so that the amount of each protein can be quantified. The amount of each protein is compared to the amount of the corresponding protein in an untreated biological sample. A difference in the amount of protein between the two samples is indicative of a toxic response to the test compound in the treated sample. Individual proteins are identified by sequencing the amino acid residues of the individual proteins and comparing these partial sequences to the polypeptides of the present invention.

In another embodiment, the toxicity of a test compound is assessed by treating a biological sample containing proteins with the test compound. Proteins from the biological sample are incubated with antibodies specific to the polypeptides of the present invention. The amount of protein recognized by the antibodies is quantified. The amount of protein in the treated biological sample is compared with the amount in an untreated biological sample. A difference in the amount of protein between the two samples is indicative of a toxic response to the test compound in the treated sample.

Microarrays may be prepared, used, and analyzed using methods known in the art (Brennan, T.M. et al. (1995) U.S. Patent No. 5,474,796; Schena, M. et al. (1996) Proc. Natl. Acad. Sci. USA 93:10614-10619; Baldeschweiler et al. (1995) PCT application WO95/251116; Shalon, D. et al. (1995) PCT application WO95/35505; Heller, R.A. et al. (1997) Proc. Natl. Acad. Sci. USA 94:2150-2155; Heller, M.J. et al. (1997) U.S. Patent No. 5,605,662). Various types of microarrays are well known and thoroughly described in Schena, M., ed. (1999; DNA Microarrays: A Practical Approach, Oxford University Press, London).

In another embodiment of the invention, nucleic acid sequences encoding NTRAN may be used to generate hybridization probes useful in mapping the naturally occurring genomic sequence. Either coding or noncoding sequences may be used, and in some instances, noncoding sequences may be preferable over coding sequences. For example, conservation of a coding sequence among members of a multi-gene family may potentially cause undesired cross hybridization during chromosomal mapping. The sequences may be mapped to a particular chromosome, to a specific region of a chromosome, or to artificial chromosome constructions, e.g., human artificial chromosomes (HACs), yeast artificial chromosomes (YACs), bacterial artificial chromosomes (BACs), bacterial P1 constructions, or single chromosome cDNA libraries (Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355; Price, C.M. (1993) Blood Rev. 7:127-134; Trask, B.J. (1991) Trends Genet. 7:149-154). Once mapped, the nucleic acid sequences may be used to develop genetic linkage maps, for example, which

correlate the inheritance of a disease state with the inheritance of a particular chromosome region or restriction fragment length polymorphism (RFLP) (Lander, E.S. and D. Botstein (1986) Proc. Natl. Acad. Sci. USA 83:7353-7357).

Fluorescent *in situ* hybridization (FISH) may be correlated with other physical and genetic map data (Heinz-Ulrich, et al. (1995) in Meyers, *supra*, pp. 965-968). Examples of genetic map data can be found in various scientific journals or at the Online Mendelian Inheritance in Man (OMIM) World Wide Web site. Correlation between the location of the gene encoding NTRAN on a physical map and a specific disorder, or a predisposition to a specific disorder, may help define the region of DNA associated with that disorder and thus may further positional cloning efforts.

In situ hybridization of chromosomal preparations and physical mapping techniques, such as linkage analysis using established chromosomal markers, may be used for extending genetic maps. Often the placement of a gene on the chromosome of another mammalian species, such as mouse, may reveal associated markers even if the exact chromosomal locus is not known. This information is valuable to investigators searching for disease genes using positional cloning or other gene discovery techniques. Once the gene or genes responsible for a disease or syndrome have been crudely localized by genetic linkage to a particular genomic region, e.g., ataxia-telangiectasia to 11q22-23, any sequences mapping to that area may represent associated or regulatory genes for further investigation (Gatti, R.A. et al. (1988) Nature 336:577-580). The nucleotide sequence of the instant invention may also be used to detect differences in the chromosomal location due to translocation, inversion, etc., among normal, carrier, or affected individuals.

In another embodiment of the invention, NTRAN, its catalytic or immunogenic fragments, or oligopeptides thereof can be used for screening libraries of compounds in any of a variety of drug screening techniques. The fragment employed in such screening may be free in solution, affixed to a solid support, borne on a cell surface, or located intracellularly. The formation of binding complexes between NTRAN and the agent being tested may be measured.

Another technique for drug screening provides for high throughput screening of compounds having suitable binding affinity to the protein of interest (Geysen, et al. (1984) PCT application WO84/03564). In this method, large numbers of different small test compounds are synthesized on a solid substrate. The test compounds are reacted with NTRAN, or fragments thereof, and washed. Bound NTRAN is then detected by methods well known in the art. Purified NTRAN can also be coated directly onto plates for use in the aforementioned drug screening techniques. Alternatively, non-neutralizing antibodies can be used to capture the peptide and immobilize it on a solid support.

In another embodiment, one may use competitive drug screening assays in which neutralizing antibodies capable of binding NTRAN specifically compete with a test compound for binding NTRAN. In this manner, antibodies can be used to detect the presence of any peptide which shares one or more antigenic determinants with NTRAN.

In additional embodiments, the nucleotide sequences which encode NTRAN may be used in any molecular biology techniques that have yet to be developed, provided the new techniques rely on properties of nucleotide sequences that are currently known, including, but not limited to, such properties as the triplet genetic code and specific base pair interactions.

Without further elaboration, it is believed that one skilled in the art can, using the preceding description, utilize the present invention to its fullest extent. The following embodiments are, therefore, to be construed as merely illustrative, and not limitative of the remainder of the disclosure in any way whatsoever.

The disclosures of all patents, applications, and publications mentioned above and below, including U.S. Ser. No. 60/340,798, U.S. Ser. No. 60/365,645, U.S. Ser. No. 60/367,662, U.S. Ser. No. 60/379,887, and U.S. Ser. No. 60/384,639, are hereby expressly incorporated by reference.

EXAMPLES

I. Construction of cDNA Libraries

Incyte cDNAs were derived from cDNA libraries described in the LIFESEQ GOLD database (Incyte Genomics, Palo Alto CA). Some tissues were homogenized and lysed in guanidinium isothiocyanate, while others were homogenized and lysed in phenol or in a suitable mixture of denaturants, such as TRIZOL (Invitrogen), a monophasic solution of phenol and guanidine isothiocyanate. The resulting lysates were centrifuged over CsCl cushions or extracted with chloroform. RNA was precipitated from the lysates with either isopropanol or sodium acetate and ethanol, or by other routine methods.

Phenol extraction and precipitation of RNA were repeated as necessary to increase RNA purity. In some cases, RNA was treated with DNase. For most libraries, poly(A)+ RNA was isolated using oligo d(T)-coupled paramagnetic particles (Promega), OLIGOTEX latex particles (QIAGEN, Chatsworth CA), or an OLIGOTEX mRNA purification kit (QIAGEN). Alternatively, RNA was isolated directly from tissue lysates using other RNA isolation kits, e.g., the POLY(A)PURE mRNA purification kit (Ambion, Austin TX).

In some cases, Stratagene was provided with RNA and constructed the corresponding cDNA

libraries. Otherwise, cDNA was synthesized and cDNA libraries were constructed with the UNIZAP vector system (Stratagene) or SUPERSCRIPT plasmid system (Invitrogen), using the recommended procedures or similar methods known in the art (Ausubel et al., *supra*, ch. 5). Reverse transcription was initiated using oligo d(T) or random primers. Synthetic oligonucleotide adapters were ligated to double stranded cDNA, and the cDNA was digested with the appropriate restriction enzyme or enzymes. For most libraries, the cDNA was size-selected (300-1000 bp) using SEPHACRYL S1000, SEPHAROSE CL2B, or SEPHAROSE CL4B column chromatography (Amersham Biosciences) or preparative agarose gel electrophoresis. cDNAs were ligated into compatible restriction enzyme sites of the polylinker of a suitable plasmid, e.g., PBLUESCRIPT plasmid (Stratagene), PSPORT1 plasmid (Invitrogen, Carlsbad CA), PCDNA2.1 plasmid (Invitrogen), PBK-CMV plasmid (Stratagene), PCR2-TOPOTA plasmid (Invitrogen), PCMV-ICIS plasmid (Stratagene), pIGEN (Incyte Genomics, Palo Alto CA), pRARE (Incyte Genomics), or pINCY (Incyte Genomics), or derivatives thereof. Recombinant plasmids were transformed into competent *E. coli* cells including XL1-Blue, XL1-BlueMRF, or SOLR from Stratagene or DH5α, DH10B, or ElectroMAX DH10B from Invitrogen.

II. Isolation of cDNA Clones

Plasmids obtained as described in Example I were recovered from host cells by *in vivo* excision using the UNIZAP vector system (Stratagene) or by cell lysis. Plasmids were purified using at least one of the following: a Magic or WIZARD Minipreps DNA purification system (Promega); an AGTC Miniprep purification kit (Edge Biosystems, Gaithersburg MD); and QIAWELL 8 Plasmid, QIAWELL 8 Plus Plasmid, QIAWELL 8 Ultra Plasmid purification systems or the R.E.A.L. PREP 96 plasmid purification kit from QIAGEN. Following precipitation, plasmids were resuspended in 0.1 ml of distilled water and stored, with or without lyophilization, at 4°C.

Alternatively, plasmid DNA was amplified from host cell lysates using direct link PCR in a high-throughput format (Rao, V.B. (1994) Anal. Biochem. 216:1-14). Host cell lysis and thermal cycling steps were carried out in a single reaction mixture. Samples were processed and stored in 384-well plates, and the concentration of amplified plasmid DNA was quantified fluorometrically using PICOGREEN dye (Molecular Probes, Eugene OR) and a FLUOROSKAN II fluorescence scanner (Labsystems Oy, Helsinki, Finland).

III. Sequencing and Analysis

Incyte cDNA recovered in plasmids as described in Example II were sequenced as follows. Sequencing reactions were processed using standard methods or high-throughput instrumentation such

as the ABI CATALYST 800 (Applied Biosystems) thermal cycler or the PTC-200 thermal cycler (MJ Research) in conjunction with the HYDRA microdispenser (Robbins Scientific) or the MICROLAB 2200 (Hamilton) liquid transfer system. cDNA sequencing reactions were prepared using reagents provided by Amersham Biosciences or supplied in ABI sequencing kits such as the ABI PRISM BIGDYE Terminator cycle sequencing ready reaction kit (Applied Biosystems). Electrophoretic separation of cDNA sequencing reactions and detection of labeled polynucleotides were carried out using the MEGABACE 1000 DNA sequencing system (Amersham Biosciences); the ABI PRISM 373 or 377 sequencing system (Applied Biosystems) in conjunction with standard ABI protocols and base calling software; or other sequence analysis systems known in the art. Reading frames within the cDNA sequences were identified using standard methods (Ausubel et al., *supra*, ch. 7). Some of the cDNA sequences were selected for extension using the techniques disclosed in Example VIII.

The polynucleotide sequences derived from Incyte cDNAs were validated by removing vector, linker, and poly(A) sequences and by masking ambiguous bases, using algorithms and programs based on BLAST, dynamic programming, and dinucleotide nearest neighbor analysis. The Incyte cDNA sequences or translations thereof were then queried against a selection of public databases such as the GenBank primate, rodent, mammalian, vertebrate, and eukaryote databases, and BLOCKS, PRINTS, DOMO, PRODOM; PROTEOME databases with sequences from *Homo sapiens*, *Rattus norvegicus*, *Mus musculus*, *Caenorhabditis elegans*, *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe*, and *Candida albicans* (Incyte Genomics, Palo Alto CA); hidden Markov model (HMM)-based protein family databases such as PFAM, INCY, and TIGRFAM (Haft, D.H. et al. (2001) Nucleic Acids Res. 29:41-43); and HMM-based protein domain databases such as SMART (Schultz, J. et al. (1998) Proc. Natl. Acad. Sci. USA 95:5857-5864; Letunic, I. et al. (2002) Nucleic Acids Res. 30:242-244). (HMM is a probabilistic approach which analyzes consensus primary structures of gene families; see, for example, Eddy, S.R. (1996) Curr. Opin. Struct. Biol. 6:361-365.) The queries were performed using programs based on BLAST, FASTA, BLIMPS, and HMMER. The Incyte cDNA sequences were assembled to produce full length polynucleotide sequences. Alternatively, GenBank cDNAs, GenBank ESTs, stitched sequences, stretched sequences, or Genscan-predicted coding sequences (see Examples IV and V) were used to extend Incyte cDNA assemblages to full length. Assembly was performed using programs based on Phred, Phrap, and Consed, and cDNA assemblages were screened for open reading frames using programs based on GeneMark, BLAST, and FASTA. The full length polynucleotide sequences were translated

to derive the corresponding full length polypeptide sequences. Alternatively, a polypeptide may begin at any of the methionine residues of the full length translated polypeptide. Full length polypeptide sequences were subsequently analyzed by querying against databases such as the GenBank protein databases (genpept), SwissProt, the PROTEOME databases, BLOCKS, PRINTS, DOMO, PRODOM, Prosite, hidden Markov model (HMM)-based protein family databases such as PFAM, INCY, and TIGRFAM; and HMM-based protein domain databases such as SMART. Full length polynucleotide sequences are also analyzed using MACDNASIS PRO software (MiraiBio, Alameda CA) and LASERGENE software (DNASTAR). Polynucleotide and polypeptide sequence alignments are generated using default parameters specified by the CLUSTAL algorithm as incorporated into the MEGALIGN multisequence alignment program (DNASTAR), which also calculates the percent identity between aligned sequences.

Table 7 summarizes the tools, programs, and algorithms used for the analysis and assembly of Incyte cDNA and full length sequences and provides applicable descriptions, references, and threshold parameters. The first column of Table 7 shows the tools, programs, and algorithms used, the second column provides brief descriptions thereof, the third column presents appropriate references, all of which are incorporated by reference herein in their entirety, and the fourth column presents, where applicable, the scores, probability values, and other parameters used to evaluate the strength of a match between two sequences (the higher the score or the lower the probability value, the greater the identity between two sequences).

The programs described above for the assembly and analysis of full length polynucleotide and polypeptide sequences were also used to identify polynucleotide sequence fragments from SEQ ID NO:23-44. Fragments from about 20 to about 4000 nucleotides which are useful in hybridization and amplification technologies are described in Table 4, column 2.

IV. Identification and Editing of Coding Sequences from Genomic DNA

Putative neurotransmission-associated proteins were initially identified by running the Genscan gene identification program against public genomic sequence databases (e.g., gbpri and gbhtg). Genscan is a general-purpose gene identification program which analyzes genomic DNA sequences from a variety of organisms (Burge, C. and S. Karlin (1997) J. Mol. Biol. 268:78-94; Burge, C. and S. Karlin (1998) Curr. Opin. Struct. Biol. 8:346-354). The program concatenates predicted exons to form an assembled cDNA sequence extending from a methionine to a stop codon. The output of Genscan is a FASTA database of polynucleotide and polypeptide sequences. The maximum range of sequence for Genscan to analyze at once was set to 30 kb. To determine which of these Genscan

predicted cDNA sequences encode neurotransmission-associated proteins, the encoded polypeptides were analyzed by querying against PFAM models for neurotransmission-associated proteins. Potential neurotransmission-associated proteins were also identified by homology to Incyte cDNA sequences that had been annotated as neurotransmission-associated proteins. These selected Genscan-predicted sequences were then compared by BLAST analysis to the genpept and gbpri public databases. Where necessary, the Genscan-predicted sequences were then edited by comparison to the top BLAST hit from genpept to correct errors in the sequence predicted by Genscan, such as extra or omitted exons. BLAST analysis was also used to find any Incyte cDNA or public cDNA coverage of the Genscan-predicted sequences, thus providing evidence for transcription. When Incyte cDNA coverage was available, this information was used to correct or confirm the Genscan predicted sequence. Full length polynucleotide sequences were obtained by assembling Genscan-predicted coding sequences with Incyte cDNA sequences and/or public cDNA sequences using the assembly process described in Example III. Alternatively, full length polynucleotide sequences were derived entirely from edited or unedited Genscan-predicted coding sequences.

V. Assembly of Genomic Sequence Data with cDNA Sequence Data

"Stitched" Sequences

Partial cDNA sequences were extended with exons predicted by the Genscan gene identification program described in Example IV. Partial cDNAs assembled as described in Example III were mapped to genomic DNA and parsed into clusters containing related cDNAs and Genscan exon predictions from one or more genomic sequences. Each cluster was analyzed using an algorithm based on graph theory and dynamic programming to integrate cDNA and genomic information, generating possible splice variants that were subsequently confirmed, edited, or extended to create a full length sequence. Sequence intervals in which the entire length of the interval was present on more than one sequence in the cluster were identified, and intervals thus identified were considered to be equivalent by transitivity. For example, if an interval was present on a cDNA and two genomic sequences, then all three intervals were considered to be equivalent. This process allows unrelated but consecutive genomic sequences to be brought together, bridged by cDNA sequence. Intervals thus identified were then "stitched" together by the stitching algorithm in the order that they appear along their parent sequences to generate the longest possible sequence, as well as sequence variants. Linkages between intervals which proceed along one type of parent sequence (cDNA to cDNA or genomic sequence to genomic sequence) were given preference over linkages which change parent type (cDNA to genomic sequence). The resultant stitched sequences were translated and compared

by BLAST analysis to the genpept and gbpri public databases. Incorrect exons predicted by Genscan were corrected by comparison to the top BLAST hit from genpept. Sequences were further extended with additional cDNA sequences, or by inspection of genomic DNA, when necessary.

"Stretched" Sequences

Partial DNA sequences were extended to full length with an algorithm based on BLAST analysis. First, partial cDNAs assembled as described in Example III were queried against public databases such as the GenBank primate, rodent, mammalian, vertebrate, and eukaryote databases using the BLAST program. The nearest GenBank protein homolog was then compared by BLAST analysis to either Incyte cDNA sequences or GenScan exon predicted sequences described in Example IV. A chimeric protein was generated by using the resultant high-scoring segment pairs (HSPs) to map the translated sequences onto the GenBank protein homolog. Insertions or deletions may occur in the chimeric protein with respect to the original GenBank protein homolog. The GenBank protein homolog, the chimeric protein, or both were used as probes to search for homologous genomic sequences from the public human genome databases. Partial DNA sequences were therefore "stretched" or extended by the addition of homologous genomic sequences. The resultant stretched sequences were examined to determine whether it contained a complete gene.

VI. Chromosomal Mapping of NTRAN Encoding Polynucleotides

The sequences which were used to assemble SEQ ID NO:23-44 were compared with sequences from the Incyte LIFESEQ database and public domain databases using BLAST and other implementations of the Smith-Waterman algorithm. Sequences from these databases that matched SEQ ID NO:23-44 were assembled into clusters of contiguous and overlapping sequences using assembly algorithms such as Phrap (Table 7). Radiation hybrid and genetic mapping data available from public resources such as the Stanford Human Genome Center (SHGC), Whitehead Institute for Genome Research (WIGR), and Généthon were used to determine if any of the clustered sequences had been previously mapped. Inclusion of a mapped sequence in a cluster resulted in the assignment of all sequences of that cluster, including its particular SEQ ID NO:, to that map location.

Map locations are represented by ranges, or intervals, of human chromosomes. The map position of an interval, in centiMorgans, is measured relative to the terminus of the chromosome's p-arm. (The centiMorgan (cM) is a unit of measurement based on recombination frequencies between chromosomal markers. On average, 1 cM is roughly equivalent to 1 megabase (Mb) of DNA in humans, although this can vary widely due to hot and cold spots of recombination.) The cM distances are based on genetic markers mapped by Généthon which provide boundaries for radiation hybrid

markers whose sequences were included in each of the clusters. Human genome maps and other resources available to the public, such as the NCBI "GeneMap'99" World Wide Web site (http://www.ncbi.nlm.nih.gov/genemap/), can be employed to determine if previously identified disease genes map within or in proximity to the intervals indicated above.

Association of a disease with a chromosomal locus can be determined by lod score. Lod score is a statistical method used to test the linkage of two or more loci within families having a genetic disease. The lod score is the logarithm to base 10 of the odds in favor of linkage. Linkage is defined as the tendency of two genes located on the same chromosome to be inherited together through meiosis (*Genetics in Medicine*, Fifth Edition, (1991) Thompson, M.W. et al., W.B. Saunders Co. Philadelphia). A lod score of +3 or greater (1000:1 odds in favor of linkage) indicates a probability of 1 in 1000 that a particular marker was found solely by chance in affected individuals, which is strong evidence that two genetic loci are linked.

One such gene implicated in PD is PARK3, which maps to 2p13 (Gasser, T. et al. (1998) Nature Genet. 18:262-265). A marker at chromosomal position D2S441 was found to have a lod score of 3.2 in the region of PARK3. This marker supported the disease association of PARK3 in the chromosomal interval from D2S134 to D2S286 (Gasser et al., supra). Markers located within chromosomal intervals D2S134 and D2S286, which map between 83.88 to 94.05 centiMorgans on the short arm of chromosome 2, were used to identify genes that map in the region between D2S134 and D2S286.

A second PD gene, implicated in early-onset recessive parkinsonism, is PARK6, located on chromosome 1 at 1p35-1p36. Several markers were obtained with lod scores greater than 3 including D1S199, D1S2732, D1S2828, D1S478, D1S2702, D1S2734, D1S2674 (Valente, E.M. et al, supra). These markers were used to determine the PD-relevant range of chromosome loci and identify sequences that map to chromosome 1 between D1S199 and D1S2885.

Restriction fragment length polymorphism (RFLP) markers shown to be near regions of DNA known as sequence-tagged sites (STS), have been mapped to NT_Contigs generated by the Human Genome Project using ePCR (Schuler, G.D. (1997) Genome Research 7: 541-550, and (1998) Trends Biotechnol. 16(11):456-9). Contigs containing regions of DNA with known disease-associated markers are therefore used to identify NTRAN sequences that map to disease-associated regions of the genome.

Polynucleotides encoding NTRAN were mapped to NT_Contigs. Contigs longer than 1Mb were broken into subcontigs of 1Mb length with overlaping sections of 100kb. A preliminary step used

an algorithm, similar to MEGABLAST, to define the mRNA sequence /masked genomic DNA contig pairings. The cDNA/genomic pairings identified by the first algorithm were confirmed, and the NTRAN polynucleotides mapped to DNA contigs, using SIM4 (Florea, L. et al. (1998) Genome Res. 8:967-74, version May 2000) which had been optimized for high throughput processing and strand assignment confidence). The SIM4 output of the mRNA sequence/genomic contig pairs was further processed to determine the correct location of the NTRAN polynucleotides on the genomic contig, as well as their strand identity.

SEQ ID NO:40 was mapped to NT_Contig GBI:NT_005428_001.7 from Genbank release February 2002, covering a 9.65 Mb region of the genome that also contains PD-associated genetic markers D2S134 and D2S286. The maximum distance between SEQ ID NO:40 and markers D2S134 and D2S286, therefore, is 9.65 Mb. Thus, SEQ ID NO:40 is in proximity with genetic markers shown to consistently associate with PD.

VII. Analysis of Polynucleotide Expression

Northern analysis is a laboratory technique used to detect the presence of a transcript of a gene and involves the hybridization of a labeled nucleotide sequence to a membrane on which RNAs from a particular cell type or tissue have been bound (Sambrook and Russell, *supra*, ch. 7; Ausubel et al., *supra*, ch. 4).

Analogous computer techniques applying BLAST were used to search for identical or related molecules in databases such as GenBank or LIFESEQ (Incyte Genomics). This analysis is much faster than multiple membrane-based hybridizations. In addition, the sensitivity of the computer search can be modified to determine whether any particular match is categorized as exact or similar. The basis of the search is the product score, which is defined as:

$$\frac{\text{BLAST Score} \times \text{Percent Identity}}{5 \times \text{minimum}\{\text{length(Seq. 1)}, \text{length(Seq. 2)}\}}$$

The product score takes into account both the degree of similarity between two sequences and the length of the sequence match. The product score is a normalized value between 0 and 100, and is calculated as follows: the BLAST score is multiplied by the percent nucleotide identity and the product is divided by (5 times the length of the shorter of the two sequences). The BLAST score is calculated by assigning a score of +5 for every base that matches in a high-scoring segment pair (HSP), and -4 for every mismatch. Two sequences may share more than one HSP (separated by gaps). If there is more than one HSP, then the pair with the highest BLAST score is used to calculate the product score. The product score represents a balance between fractional overlap and quality in a

BLAST alignment. For example, a product score of 100 is produced only for 100% identity over the entire length of the shorter of the two sequences being compared. A product score of 70 is produced either by 100% identity and 70% overlap at one end, or by 88% identity and 100% overlap at the other. A product score of 50 is produced either by 100% identity and 50% overlap at one end, or 79% identity and 100% overlap.

Alternatively, polynucleotides encoding NTRAN are analyzed with respect to the tissue sources from which they were derived. For example, some full length sequences are assembled, at least in part, with overlapping Incyte cDNA sequences (see Example III). Each cDNA sequence is derived from a cDNA library constructed from a human tissue. Each human tissue is classified into one of the following organ/tissue categories: cardiovascular system; connective tissue; digestive system; embryonic structures; endocrine system; exocrine glands; genitalia, female; genitalia, male; germ cells; hemic and immune system; liver; musculoskeletal system; nervous system; pancreas; respiratory system; sense organs; skin; stomatognathic system; unclassified/mixed; or urinary tract. The number of libraries in each category is counted and divided by the total number of libraries across all categories. Similarly, each human tissue is classified into one of the following disease/condition categories: cancer, cell line, developmental, inflammation, neurological, trauma, cardiovascular, pooled, and other, and the number of libraries in each category is counted and divided by the total number of libraries across all categories. The resulting percentages reflect the tissue- and disease-specific expression of cDNA encoding NTRAN. cDNA sequences and cDNA library/tissue information are found in the LIFESEQ GOLD database (Incyte Genomics, Palo Alto CA).

VIII. Extension of NTRAN Encoding Polynucleotides

Full length polynucleotides are produced by extension of an appropriate fragment of the full length molecule using oligonucleotide primers designed from this fragment. One primer was synthesized to initiate 5' extension of the known fragment, and the other primer was synthesized to initiate 3' extension of the known fragment. The initial primers were designed using OLIGO 4.06 software (National Biosciences), or another appropriate program, to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the target sequence at temperatures of about 68°C to about 72°C. Any stretch of nucleotides which would result in hairpin structures and primer-primer dimerizations was avoided.

Selected human cDNA libraries were used to extend the sequence. If more than one extension was necessary or desired, additional or nested sets of primers were designed.

High fidelity amplification was obtained by PCR using methods well known in the art. PCR

was performed in 96-well plates using the PTC-200 thermal cycler (MJ Research, Inc.). The reaction mix contained DNA template, 200 nmol of each primer, reaction buffer containing Mg^{2+}, $(NH_4)_2SO_4$, and 2-mercaptoethanol, Taq DNA polymerase (Amersham Biosciences), ELONGASE enzyme (Invitrogen), and Pfu DNA polymerase (Stratagene), with the following parameters for primer pair PCI A and PCI B: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 60°C, 1 min; Step 4: 68°C, 2 min; Step 5: Steps 2, 3, and 4 repeated 20 times; Step 6: 68°C, 5 min; Step 7: storage at 4°C. In the alternative, the parameters for primer pair T7 and SK+ were as follows: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 57°C, 1 min; Step 4: 68°C, 2 min; Step 5: Steps 2, 3, and 4 repeated 20 times; Step 6: 68°C, 5 min; Step 7: storage at 4°C.

The concentration of DNA in each well was determined by dispensing 100 μl PICOGREEN quantitation reagent (0.25% (v/v) PICOGREEN; Molecular Probes, Eugene OR) dissolved in 1X TE and 0.5 μl of undiluted PCR product into each well of an opaque fluorimeter plate (Corning Costar, Acton MA), allowing the DNA to bind to the reagent. The plate was scanned in a Fluoroskan II (Labsystems Oy, Helsinki, Finland) to measure the fluorescence of the sample and to quantify the concentration of DNA. A 5 μl to 10 μl aliquot of the reaction mixture was analyzed by electrophoresis on a 1 % agarose gel to determine which reactions were successful in extending the sequence.

The extended nucleotides were desalted and concentrated, transferred to 384-well plates, digested with CviJI cholera virus endonuclease (Molecular Biology Research, Madison WI), and sonicated or sheared prior to religation into pUC 18 vector (Amersham Biosciences). For shotgun sequencing, the digested nucleotides were separated on low concentration (0.6 to 0.8%) agarose gels, fragments were excised, and agar digested with Agar ACE (Promega). Extended clones were religated using T4 ligase (New England Biolabs, Beverly MA) into pUC 18 vector (Amersham Biosciences), treated with Pfu DNA polymerase (Stratagene) to fill-in restriction site overhangs, and transfected into competent *E. coli* cells. Transformed cells were selected on antibiotic-containing media, and individual colonies were picked and cultured overnight at 37°C in 384-well plates in LB/2x carb liquid media.

The cells were lysed, and DNA was amplified by PCR using Taq DNA polymerase (Amersham Biosciences) and Pfu DNA polymerase (Stratagene) with the following parameters: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 60°C, 1 min; Step 4: 72°C, 2 min; Step 5: steps 2, 3, and 4 repeated 29 times; Step 6: 72°C, 5 min; Step 7: storage at 4°C. DNA was quantified by PICOGREEN reagent (Molecular Probes) as described above. Samples with low DNA recoveries

were reamplified using the same conditions as described above. Samples were diluted with 20% dimethysulfoxide (1:2, v/v), and sequenced using DYENAMIC energy transfer sequencing primers and the DYENAMIC DIRECT kit (Amersham Biosciences) or the ABI PRISM BIGDYE Terminator cycle sequencing ready reaction kit (Applied Biosystems).

In like manner, full length polynucleotides are verified using the above procedure or are used to obtain 5' regulatory sequences using the above procedure along with oligonucleotides designed for such extension, and an appropriate genomic library.

IX. Identification of Single Nucleotide Polymorphisms in NTRAN Encoding Polynucleotides

Common DNA sequence variants known as single nucleotide polymorphisms (SNPs) were identified in SEQ ID NO:23-44 using the LIFESEQ database (Incyte Genomics). Sequences from the same gene were clustered together and assembled as described in Example III, allowing the identification of all sequence variants in the gene. An algorithm consisting of a series of filters was used to distinguish SNPs from other sequence variants. Preliminary filters removed the majority of basecall errors by requiring a minimum Phred quality score of 15, and removed sequence alignment errors and errors resulting from improper trimming of vector sequences, chimeras, and splice variants. An automated procedure of advanced chromosome analysis analysed the original chromatogram files in the vicinity of the putative SNP. Clone error filters used statistically generated algorithms to identify errors introduced during laboratory processing, such as those caused by reverse transcriptase, polymerase, or somatic mutation. Clustering error filters used statistically generated algorithms to identify errors resulting from clustering of close homologs or pseudogenes, or due to contamination by non-human sequences. A final set of filters removed duplicates and SNPs found in immunoglobulins or T-cell receptors.

Certain SNPs were selected for further characterization by mass spectrometry using the high throughput MASSARRAY system (Sequenom, Inc.) to analyze allele frequencies at the SNP sites in four different human populations. The Caucasian population comprised 92 individuals (46 male, 46 female), including 83 from Utah, four French, three Venezualan, and two Amish individuals. The African population comprised 194 individuals (97 male, 97 female), all African Americans. The Hispanic population comprised 324 individuals (162 male, 162 female), all Mexican Hispanic. The Asian population comprised 126 individuals (64 male, 62 female) with a reported parental breakdown of 43% Chinese, 31% Japanese, 13% Korean, 5% Vietnamese, and 8% other Asian. Allele frequencies were first analyzed in the Caucasian population; in some cases those SNPs which showed

no allelic variance in this population were not further tested in the other three populations.

X. Labeling and Use of Individual Hybridization Probes

Hybridization probes derived from SEQ ID NO:23-44 are employed to screen cDNAs, genomic DNAs, or mRNAs. Although the labeling of oligonucleotides, consisting of about 20 base pairs, is specifically described, essentially the same procedure is used with larger nucleotide fragments. Oligonucleotides are designed using state-of-the-art software such as OLIGO 4.06 software (National Biosciences) and labeled by combining 50 pmol of each oligomer, 250 μCi of [γ-^{32}P] adenosine triphosphate (Amersham Biosciences), and T4 polynucleotide kinase (DuPont NEN, Boston MA). The labeled oligonucleotides are substantially purified using a SEPHADEX G-25 superfine size exclusion dextran bead column (Amersham Biosciences). An aliquot containing 10^7 counts per minute of the labeled probe is used in a typical membrane-based hybridization analysis of human genomic DNA digested with one of the following endonucleases: Ase I, Bgl II, Eco RI, Pst I, Xba I, or Pvu II (DuPont NEN).

The DNA from each digest is fractionated on a 0.7% agarose gel and transferred to nylon membranes (Nytran Plus, Schleicher & Schuell, Durham NH). Hybridization is carried out for 16 hours at 40°C. To remove nonspecific signals, blots are sequentially washed at room temperature under conditions of up to, for example, 0.1 x saline sodium citrate and 0.5% sodium dodecyl sulfate. Hybridization patterns are visualized using autoradiography or an alternative imaging means and compared.

XI. Microarrays

The linkage or synthesis of array elements upon a microarray can be achieved utilizing photolithography, piezoelectric printing (ink-jet printing; see, e.g., Baldeschweiler et al., *supra*), mechanical microspotting technologies, and derivatives thereof. The substrate in each of the aforementioned technologies should be uniform and solid with a non-porous surface (Schena, M., ed. (1999) DNA Microarrays: A Practical Approach, Oxford University Press, London). Suggested substrates include silicon, silica, glass slides, glass chips, and silicon wafers. Alternatively, a procedure analogous to a dot or slot blot may also be used to arrange and link elements to the surface of a substrate using thermal, UV, chemical, or mechanical bonding procedures. A typical array may be produced using available methods and machines well known to those of ordinary skill in the art and may contain any appropriate number of elements (Schena, M. et al. (1995) Science 270:467-470; Shalon, D. et al. (1996) Genome Res. 6:639-645; Marshall, A. and J. Hodgson (1998) Nat. Biotechnol. 16:27-31).

Full length cDNAs, Expressed Sequence Tags (ESTs), or fragments or oligomers thereof may comprise the elements of the microarray. Fragments or oligomers suitable for hybridization can be selected using software well known in the art such as LASERGENE software (DNASTAR). The array elements are hybridized with polynucleotides in a biological sample. The polynucleotides in the biological sample are conjugated to a fluorescent label or other molecular tag for ease of detection. After hybridization, nonhybridized nucleotides from the biological sample are removed, and a fluorescence scanner is used to detect hybridization at each array element. Alternatively, laser desorbtion and mass spectrometry may be used for detection of hybridization. The degree of complementarity and the relative abundance of each polynucleotide which hybridizes to an element on the microarray may be assessed. In one embodiment, microarray preparation and usage is described in detail below.

Tissue or Cell Sample Preparation

Total RNA is isolated from tissue samples using the guanidinium thiocyanate method and poly(A)$^+$ RNA is purified using the oligo-(dT) cellulose method. Each poly(A)$^+$ RNA sample is reverse transcribed using MMLV reverse-transcriptase, 0.05 pg/μl oligo-(dT) primer (21mer), 1X first strand buffer, 0.03 units/μl RNase inhibitor, 500 μM dATP, 500 μM dGTP, 500 μM dTTP, 40 μM dCTP, 40 μM dCTP-Cy3 (BDS) or dCTP-Cy5 (Amersham Biosciences). The reverse transcription reaction is performed in a 25 ml volume containing 200 ng poly(A)$^+$ RNA with GEMBRIGHT kits (Incyte Genomics). Specific control poly(A)$^+$ RNAs are synthesized by *in vitro* transcription from non-coding yeast genomic DNA. After incubation at 37° C for 2 hr, each reaction sample (one with Cy3 and another with Cy5 labeling) is treated with 2.5 ml of 0.5M sodium hydroxide and incubated for 20 minutes at 85° C to the stop the reaction and degrade the RNA. Samples are purified using two successive CHROMA SPIN 30 gel filtration spin columns (Clontech, Palo Alto CA) and after combining, both reaction samples are ethanol precipitated using 1 ml of glycogen (1 mg/ml), 60 ml sodium acetate, and 300 ml of 100% ethanol. The sample is then dried to completion using a SpeedVAC (Savant Instruments Inc., Holbrook NY) and resuspended in 14 μl 5X SSC/0.2% SDS.

Microarray Preparation

Sequences of the present invention are used to generate array elements. Each array element is amplified from bacterial cells containing vectors with cloned cDNA inserts. PCR amplification uses primers complementary to the vector sequences flanking the cDNA insert. Array elements are amplified in thirty cycles of PCR from an initial quantity of 1-2 ng to a final quantity greater than 5 μg. Amplified array elements are then purified using SEPHACRYL-400 (Amersham Biosciences).

Purified array elements are immobilized on polymer-coated glass slides. Glass microscope slides (Corning) are cleaned by ultrasound in 0.1% SDS and acetone, with extensive distilled water washes between and after treatments. Glass slides are etched in 4% hydrofluoric acid (VWR Scientific Products Corporation (VWR), West Chester PA), washed extensively in distilled water, and coated with 0.05% aminopropyl silane (Sigma) in 95% ethanol. Coated slides are cured in a 110°C oven.

Array elements are applied to the coated glass substrate using a procedure described in U.S. Patent No. 5,807,522, incorporated herein by reference. 1 μl of the array element DNA, at an average concentration of 100 ng/μl, is loaded into the open capillary printing element by a high-speed robotic apparatus. The apparatus then deposits about 5 nl of array element sample per slide.

Microarrays are UV-crosslinked using a STRATALINKER UV-crosslinker (Stratagene). Microarrays are washed at room temperature once in 0.2% SDS and three times in distilled water. Non-specific binding sites are blocked by incubation of microarrays in 0.2% casein in phosphate buffered saline (PBS) (Tropix, Inc., Bedford MA) for 30 minutes at 60° C followed by washes in 0.2% SDS and distilled water as before.

Hybridization

Hybridization reactions contain 9 μl of sample mixture consisting of 0.2 μg each of Cy3 and Cy5 labeled cDNA synthesis products in 5X SSC, 0.2% SDS hybridization buffer. The sample mixture is heated to 65° C for 5 minutes and is aliquoted onto the microarray surface and covered with an 1.8 cm^2 coverslip. The arrays are transferred to a waterproof chamber having a cavity just slightly larger than a microscope slide. The chamber is kept at 100% humidity internally by the addition of 140 μl of 5X SSC in a corner of the chamber. The chamber containing the arrays is incubated for about 6.5 hours at 60° C. The arrays are washed for 10 min at 45° C in a first wash buffer (1X SSC, 0.1% SDS), three times for 10 minutes each at 45° C in a second wash buffer (0.1X SSC), and dried.

Detection

Reporter-labeled hybridization complexes are detected with a microscope equipped with an Innova 70 mixed gas 10 W laser (Coherent, Inc., Santa Clara CA) capable of generating spectral lines at 488 nm for excitation of Cy3 and at 632 nm for excitation of Cy5. The excitation laser light is focused on the array using a 20X microscope objective (Nikon, Inc., Melville NY). The slide containing the array is placed on a computer-controlled X-Y stage on the microscope and raster-scanned past the objective. The 1.8 cm x 1.8 cm array used in the present example is scanned with a resolution of 20 micrometers.

In two separate scans, a mixed gas multiline laser excites the two fluorophores sequentially. Emitted light is split, based on wavelength, into two photomultiplier tube detectors (PMT R1477, Hamamatsu Photonics Systems, Bridgewater NJ) corresponding to the two fluorophores. Appropriate filters positioned between the array and the photomultiplier tubes are used to filter the signals. The emission maxima of the fluorophores used are 565 nm for Cy3 and 650 nm for Cy5. Each array is typically scanned twice, one scan per fluorophore using the appropriate filters at the laser source, although the apparatus is capable of recording the spectra from both fluorophores simultaneously.

The sensitivity of the scans is typically calibrated using the signal intensity generated by a cDNA control species added to the sample mixture at a known concentration. A specific location on the array contains a complementary DNA sequence, allowing the intensity of the signal at that location to be correlated with a weight ratio of hybridizing species of 1:100,000. When two samples from different sources (e.g., representing test and control cells), each labeled with a different fluorophore, are hybridized to a single array for the purpose of identifying genes that are differentially expressed, the calibration is done by labeling samples of the calibrating cDNA with the two fluorophores and adding identical amounts of each to the hybridization mixture.

The output of the photomultiplier tube is digitized using a 12-bit RTI-835H analog-to-digital (A/D) conversion board (Analog Devices, Inc., Norwood MA) installed in an IBM-compatible PC computer. The digitized data are displayed as an image where the signal intensity is mapped using a linear 20-color transformation to a pseudocolor scale ranging from blue (low signal) to red (high signal). The data is also analyzed quantitatively. Where two different fluorophores are excited and measured simultaneously, the data are first corrected for optical crosstalk (due to overlapping emission spectra) between the fluorophores using each fluorophore's emission spectrum.

A grid is superimposed over the fluorescence signal image such that the signal from each spot is centered in each element of the grid. The fluorescence signal within each element is then integrated to obtain a numerical value corresponding to the average intensity of the signal. The software used for signal analysis is the GEMTOOLS gene expression analysis program (Incyte Genomics). Array elements that exhibit at least about a two-fold change in expression, a signal-to-background ratio of at least about 2.5, and an element spot size of at least about 40%, are considered to be differentially expressed.

Expression

For example, SEQ ID NO:24 showed differential expression in lung squamous carcinoma tissues versus normal lung tissues as determined by microarray analysis. The expression of SEQ ID

NO:24 was decreased by at least two fold in lung squamous carcinoma tissues relative to grossly uninvolved normal lung tissue from the same donors. Therefore, SEQ ID NO:24 is useful as a diagnostic marker or as a potential therapeutic target for lung cancer.

SEQ ID NO:24 also showed differential expression in ovarian adenocarcinoma tissues versus normal ovarian tissues as determined by microarray analysis. The expression of SEQ ID NO:24 was decreased by at least two fold in ovarian adenocarcinoma tissues relative to grossly uninvolved normal ovarian tissue from the same donor. Therefore, SEQ ID NO:24 is useful as a diagnostic marker or as a potential therapeutic target for ovarian cancer.

Furthermore, SEQ ID NO:24 showed region-specific gene expression in the human brain as determined by microarray analysis. The expression of SEQ ID NO:24 was decreased by at least two fold in the cervical spinal cord relative to pooled brain tissues which were constituted from the major regions of the brain from two male brains; a 47 year old and a 48 year old. The tissue from the cervical spinal cord was isolated from a 47 year old male, the same 47 year old donor as in the pooled sample. Therefore, SEQ ID NO:24 serves as a useful biomarker for human brains, specifically for the cervical spinal cord.

In an alternative example, the expression of SEQ ID NO:25, as determined by microarray analysis, was increased by at least two fold in sigmoid colon tissues relative to normal sigmoid colon tissues. The sigmoid colon tumor tissue which originated from a metastatic gastric sarcoma (stromal tumor) was harvested from a 48 year old female donor. The normal sigmoid colon tissue was harvested from grossly uninvolved sigmoid colon tissue of the same donor. Therefore, SEQ ID NO:25 is useful as a diagnostic marker or as a potential therapeutic target for colon cancer.

SEQ ID NO:25 also showed decreased expression in tissue affected by lung adenocarcinoma versus normal lung tissue as determined by microarray analysis. A sample of right lung tissue that showed moderately differentiated adenocarcinoma was compared to grossly uninvolved lung tissue from the same donor (Huntsman Cancer Institute, Salt Lake City, UT). Therefore, SEQ ID NO:25 is useful in diagnostic assays for and monitoring treatment of lung cancer.

Furthermore, the expression of SEQ ID NO:25, as determined by microarray analysis, was increased by at least two fold in Tangier disease-derived fibroblasts relative to normal fibroblasts. Both types of cells were cultured in the presence of cholesterol and compared with the same cell type in the absence of cholesterol. The human fibroblasts were obtained from skin transplants from both normal subjects and two patients with homozygous Tangier disease. Cell lines were immortalized by transfection with human papillomavirus 16 genes E6 and E7 and a neomycin resistance selectable

marker. TD derived cells are deficient in an assay of apoA-I mediated tritiated cholesterol efflux. Therefore, SEQ ID NO:25 is useful in diagnostic assays for and monitoring treatment of Tangier disease.

In an alternative example, SEQ ID NO:30 showed decreased expression in lung tissue affected by squamous cell adenocarcinoma versus normal lung tissue as determined by microarray analysis. Grossly uninvolved lung tissue with no visible abnormalities, from a 73 year-old male, was compared to lung squamous cell adenocarcinoma tissue from the same donor (Roy Castle International Centre for Lung Cancer Research, Liverpool, UK). Therefore, SEQ ID NO:30 is useful in monitoring treatment of, and diagnostic assays for, lung cancer and other cell proliferative disorders.

As another example, SEQ ID NO:30 showed decreased expression in tissue affected by ovarian tumor versus normal ovary tissue as determined by microarray analysis. A normal ovary from a 79 year-old female donor was compared to an ovarian tumor from the same donor (Huntsman Cancer Institute, Salt Lake City, UT). Therefore, SEQ ID NO:30 is useful in monitoring treatment of, and diagnostic assays for, ovarian cancer and other cell proliferative disorders.

In an alternative example, SEQ ID NO:36 showed decreased expression in colon adenocarcinoma tissue versus grossly uninvolved colon tissue. Gene expression profiles were obtained by comparing normal colon tissue to colon tumor tissue from the same donor. Therefore, SEQ ID NO:36 is useful in monitoring treatment of, and diagnostic assays for, colon cancer and other cell proliferative disorders.

In an alternative example, SEQ ID NO:38 showed differential expression, as determined by microarray analysis. Array elements that exhibited about at least a two-fold change in expression and a signal intensity over 250 units, a signal-to-background ratio of a least 2.5, and an element spot size of at least 40% were identified as differentially expressed using the GEMTOOLS program (Incyte Genomics).

For example, the expression of cDNA from grossly uninvolved lung tissue from a 66 year-old male was compared to lung squamous cell carcinoma tissue from the same donor (Roy Castle International Centre for Lung Cancer Research, Liverpool, UK). The expression of SEQ ID NO:38 showed at least a two-fold increase in expression in the lung tumor tissue when compared to normal lung tissue expression levels from the same donor. Thus, SEQ ID NO:38 is useful for monitoring progression of, and diagnostic assays for, lung cancers.

In another example, specific dissected brain regions from a normal 61-year-old female were compared to dissected regions from the brain of a 79-year-old-female with severe Alzheimer's

disease (AD). The diagnosis of normal or severe AD was established by a certified neuropathologist based on microscopic examination of multiple sections throughout the brain. The expression of SEQ ID NO:38 showed at least a two-fold decrease in expression in the amygdala and anterior hippocampus brain regions when compared to the same brain regions from the normal 61-year-old female. Therefore, SEQ ID NO:38 is also useful for monitoring progression of, and diagnostic assays for, Alzheimer's disease.

In another example, SEQ ID NO:41 showed tissue-specific expression. RNA samples isolated from a variety of normal human tissues were compared to a common reference sample. Tissues contributing to the reference sample were selected for their ability to provide a complete distribution of RNA in the human body and include brain (4%), heart (7%), kidney (3%), lung (8%), placenta (46%), small intestine (9%), spleen (3%), stomach (6%), testis (9%), and uterus (5%). The normal tissues assayed were obtained from at least three different donors. RNA from each donor was separately isolated and individually hybridized to the microarray. Since these hybridization experiments were conducted using a common reference sample, differential expression values are directly comparable from one tissue to another. The expression of SEQ ID NO:41 was increased by at least two-fold in brain (amygdaloid body, occipital cortex, and parietal cortex) tissues as compared to the reference sample. Therefore, SEQ ID NO:41 can be used as a marker for brain tissues.

XII. Complementary Polynucleotides

Sequences complementary to the NTRAN-encoding sequences, or any parts thereof, are used to detect, decrease, or inhibit expression of naturally occurring NTRAN. Although use of oligonucleotides comprising from about 15 to 30 base pairs is described, essentially the same procedure is used with smaller or with larger sequence fragments. Appropriate oligonucleotides are designed using OLIGO 4.06 software (National Biosciences) and the coding sequence of NTRAN. To inhibit transcription, a complementary oligonucleotide is designed from the most unique 5' sequence and used to prevent promoter binding to the coding sequence. To inhibit translation, a complementary oligonucleotide is designed to prevent ribosomal binding to the NTRAN-encoding transcript.

XIII. Expression of NTRAN

Expression and purification of NTRAN is achieved using bacterial or virus-based expression systems. For expression of NTRAN in bacteria, cDNA is subcloned into an appropriate vector containing an antibiotic resistance gene and an inducible promoter that directs high levels of cDNA transcription. Examples of such promoters include, but are not limited to, the *trp-lac* (*tac*) hybrid promoter and the T5 or T7 bacteriophage promoter in conjunction with the *lac* operator regulatory

element. Recombinant vectors are transformed into suitable bacterial hosts, e.g., BL21(DE3). Antibiotic resistant bacteria express NTRAN upon induction with isopropyl beta-D-thiogalactopyranoside (IPTG). Expression of NTRAN in eukaryotic cells is achieved by infecting insect or mammalian cell lines with recombinant *Autographica californica* nuclear polyhedrosis virus (AcMNPV), commonly known as baculovirus. The nonessential polyhedrin gene of baculovirus is replaced with cDNA encoding NTRAN by either homologous recombination or bacterial-mediated transposition involving transfer plasmid intermediates. Viral infectivity is maintained and the strong polyhedrin promoter drives high levels of cDNA transcription. Recombinant baculovirus is used to infect *Spodoptera frugiperda* (Sf9) insect cells in most cases, or human hepatocytes, in some cases. Infection of the latter requires additional genetic modifications to baculovirus (Engelhard, E.K. et al. (1994) Proc. Natl. Acad. Sci. USA 91:3224-3227; Sandig, V. et al. (1996) Hum. Gene Ther. 7:1937-1945).

In most expression systems, NTRAN is synthesized as a fusion protein with, e.g., glutathione S-transferase (GST) or a peptide epitope tag, such as FLAG or 6-His, permitting rapid, single-step, affinity-based purification of recombinant fusion protein from crude cell lysates. GST, a 26-kilodalton enzyme from *Schistosoma japonicum*, enables the purification of fusion proteins on immobilized glutathione under conditions that maintain protein activity and antigenicity (Amersham Biosciences). Following purification, the GST moiety can be proteolytically cleaved from NTRAN at specifically engineered sites. FLAG, an 8-amino acid peptide, enables immunoaffinity purification using commercially available monoclonal and polyclonal anti-FLAG antibodies (Eastman Kodak). 6-His, a stretch of six consecutive histidine residues, enables purification on metal-chelate resins (QIAGEN). Methods for protein expression and purification are discussed in Ausubel et al. (*supra*, ch. 10 and 16). Purified NTRAN obtained by these methods can be used directly in the assays shown in Examples XVII and XVIII, where applicable.

XIV. Functional Assays

NTRAN function is assessed by expressing the sequences encoding NTRAN at physiologically elevated levels in mammalian cell culture systems. cDNA is subcloned into a mammalian expression vector containing a strong promoter that drives high levels of cDNA expression. Vectors of choice include PCMV SPORT plasmid (Invitrogen, Carlsbad CA) and PCR3.1 plasmid (Invitrogen), both of which contain the cytomegalovirus promoter. 5-10 μg of recombinant vector are transiently transfected into a human cell line, for example, an endothelial or hematopoietic cell line, using either liposome formulations or electroporation. 1-2 μg of an additional

plasmid containing sequences encoding a marker protein are co-transfected. Expression of a marker protein provides a means to distinguish transfected cells from nontransfected cells and is a reliable predictor of cDNA expression from the recombinant vector. Marker proteins of choice include, e.g., Green Fluorescent Protein (GFP; Clontech), CD64, or a CD64-GFP fusion protein. Flow cytometry (FCM), an automated, laser optics-based technique, is used to identify transfected cells expressing GFP or CD64-GFP and to evaluate the apoptotic state of the cells and other cellular properties. FCM detects and quantifies the uptake of fluorescent molecules that diagnose events preceding or coincident with cell death. These events include changes in nuclear DNA content as measured by staining of DNA with propidium iodide; changes in cell size and granularity as measured by forward light scatter and 90 degree side light scatter; down-regulation of DNA synthesis as measured by decrease in bromodeoxyuridine uptake; alterations in expression of cell surface and intracellular proteins as measured by reactivity with specific antibodies; and alterations in plasma membrane composition as measured by the binding of fluorescein-conjugated Annexin V protein to the cell surface. Methods in flow cytometry are discussed in Ormerod, M.G. (1994; Flow Cytometry, Oxford, New York NY).

The influence of NTRAN on gene expression can be assessed using highly purified populations of cells transfected with sequences encoding NTRAN and either CD64 or CD64-GFP. CD64 and CD64-GFP are expressed on the surface of transfected cells and bind to conserved regions of human immunoglobulin G (IgG). Transfected cells are efficiently separated from nontransfected cells using magnetic beads coated with either human IgG or antibody against CD64 (DYNAL, Lake Success NY). mRNA can be purified from the cells using methods well known by those of skill in the art. Expression of mRNA encoding NTRAN and other genes of interest can be analyzed by northern analysis or microarray techniques.

XV. Production of NTRAN Specific Antibodies

NTRAN substantially purified using polyacrylamide gel electrophoresis (PAGE; see, e.g., Harrington, M.G. (1990) Methods Enzymol. 182:488-495), or other purification techniques, is used to immunize animals (e.g., rabbits, mice, etc.) and to produce antibodies using standard protocols.

Alternatively, the NTRAN amino acid sequence is analyzed using LASERGENE software (DNASTAR) to determine regions of high immunogenicity, and a corresponding oligopeptide is synthesized and used to raise antibodies by means known to those of skill in the art. Methods for selection of appropriate epitopes, such as those near the C-terminus or in hydrophilic regions are well described in the art (Ausubel et al., *supra*, ch. 11).

Typically, oligopeptides of about 15 residues in length are synthesized using an ABI 431A peptide synthesizer (Applied Biosystems) using FMOC chemistry and coupled to KLH (Sigma-Aldrich, St. Louis MO) by reaction with N-maleimidobenzoyl-N-hydroxysuccinimide ester (MBS) to increase immunogenicity (Ausubel et al., *supra*). Rabbits are immunized with the oligopeptide-KLH complex in complete Freund's adjuvant. Resulting antisera are tested for antipeptide and anti-NTRAN activity by, for example, binding the peptide or NTRAN to a substrate, blocking with 1% BSA, reacting with rabbit antisera, washing, and reacting with radio-iodinated goat anti-rabbit IgG.

XVI. Purification of Naturally Occurring NTRAN Using Specific Antibodies

Naturally occurring or recombinant NTRAN is substantially purified by immunoaffinity chromatography using antibodies specific for NTRAN. An immunoaffinity column is constructed by covalently coupling anti-NTRAN antibody to an activated chromatographic resin, such as CNBr-activated SEPHAROSE (Amersham Biosciences). After the coupling, the resin is blocked and washed according to the manufacturer's instructions.

Media containing NTRAN are passed over the immunoaffinity column, and the column is washed under conditions that allow the preferential absorbance of NTRAN (e.g., high ionic strength buffers in the presence of detergent). The column is eluted under conditions that disrupt antibody/NTRAN binding (e.g., a buffer of pH 2 to pH 3, or a high concentration of a chaotrope, such as urea or thiocyanate ion), and NTRAN is collected.

XVII. Identification of Molecules Which Interact with NTRAN

NTRAN, or biologically active fragments thereof, are labeled with ^{125}I Bolton-Hunter reagent (Bolton, A.E. and W.M. Hunter (1973) Biochem. J. 133:529-539). Candidate molecules previously arrayed in the wells of a multi-well plate are incubated with the labeled NTRAN, washed, and any wells with labeled NTRAN complex are assayed. Data obtained using different concentrations of NTRAN are used to calculate values for the number, affinity, and association of NTRAN with the candidate molecules.

Alternatively, molecules interacting with NTRAN are analyzed using the yeast two-hybrid system as described in Fields, S. and O. Song (1989; Nature 340:245-246), or using commercially available kits based on the two-hybrid system, such as the MATCHMAKER system (Clontech).

NTRAN may also be used in the PATHCALLING process (CuraGen Corp., New Haven CT) which employs the yeast two-hybrid system in a high-throughput manner to determine all interactions between the proteins encoded by two large libraries of genes (Nandabalan, K. et al. (2000) U.S. Patent No. 6,057,101).

XVIII. Demonstration of NTRAN Activity

An assay for NTRAN activity measures the expression of NTRAN on the cell surface. cDNA encoding NTRAN is transfected into an appropriate mammalian cell line. Cell surface proteins are labeled with biotin as described (de la Fuente, M.A. et al. (1997) Blood 90:2398-2405). Immunoprecipitations are performed using NTRAN-specific antibodies, and immunoprecipitated samples are analyzed using sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) and immunoblotting techniques. The ratio of labeled immunoprecipitant to unlabeled immunoprecipitant is proportional to the amount of NTRAN expressed on the cell surface.

In the alternative, an assay for NTRAN activity is based on a prototypical assay for ligand/receptor-mediated modulation of cell proliferation. This assay measures the rate of DNA synthesis in Swiss mouse 3T3 cells. A plasmid containing polynucleotides encoding NTRAN is added to quiescent 3T3 cultured cells using transfection methods well known in the art. The transiently transfected cells are then incubated in the presence of [^{3}H]thymidine, a radioactive DNA precursor molecule. Varying amounts of NTRAN ligand are then added to the cultured cells. Incorporation of [^{3}H]thymidine into acid-precipitable DNA is measured over an appropriate time interval using a radioisotope counter, and the amount incorporated is directly proportional to the amount of newly synthesized DNA. A linear dose-response curve over at least a hundred-fold NTRAN ligand concentration range is indicative of receptor activity. One unit of activity per milliliter is defined as the concentration of NTRAN producing a 50% response level, where 100% represents maximal incorporation of [^{3}H]thymidine into acid-precipitable DNA (McKay, I. and I. Leigh, eds. (1993) Growth Factors: A Practical Approach, Oxford University Press, New York NY, p. 73.)

In a further alternative, the assay for NTRAN activity is based upon the ability of GPCR family proteins to modulate G protein-activated second messenger signal transduction pathways (e.g., cAMP; Gaudin, P. et al. (1998) J. Biol. Chem. 273:4990-4996). A plasmid encoding full length NTRAN is transfected into a mammalian cell line (e.g., Chinese hamster ovary (CHO) or human embryonic kidney (HEK-293) cell lines) using methods well-known in the art. Transfected cells are grown in 12-well trays in culture medium for 48 hours, then the culture medium is discarded, and the attached cells are gently washed with PBS. The cells are then incubated in culture medium with or without ligand for 30 minutes, then the medium is removed and cells lysed by treatment with 1 M perchloric acid. The cAMP levels in the lysate are measured by radioimmunoassay using methods well-known in the art. Changes in the levels of cAMP in the lysate from cells exposed to ligand compared to those without ligand are proportional to the amount of NTRAN present in the transfected

cells.

To measure changes in inositol phosphate levels, the cells are grown in 24-well plates containing $1x10^5$ cells/well and incubated with inositol-free media and [^{3}H]myoinositol, 2 mCi/well, for 48 hr. The culture medium is removed, and the cells washed with buffer containing 10 mM LiCl followed by addition of ligand. The reaction is stopped by addition of perchloric acid. Inositol phosphates are extracted and separated on Dowex AG1-X8 (Bio-Rad) anion exchange resin, and the total labeled inositol phosphates counted by liquid scintillation. Changes in the levels of labeled inositol phosphate from cells exposed to ligand compared to those without ligand are proportional to the amount of NTRAN present in the transfected cells.

In a further alternative, the ion conductance capacity of NTRAN is demonstrated using an electrophysiological assay. NTRAN is expressed by transforming a mammalian cell line such as COS7, HeLa or CHO with a eukaryotic expression vector encoding NTRAN. Eukaryotic expression vectors are commercially available, and the techniques to introduce them into cells are well known to those skilled in the art. A small amount of a second plasmid, which expresses any one of a number of marker genes such as β-galactosidase, is co-transformed into the cells in order to allow rapid identification of those cells which have taken up and expressed the foreign DNA. The cells are incubated for 48-72 hours after transformation under conditions appropriate for the cell line to allow expression and accumulation of NTRAN and β-galactosidase. Transformed cells expressing β-galactosidase are stained blue when a suitable colorimetric substrate is added to the culture media under conditions that are well known in the art. Stained cells are tested for differences in membrane conductance due to various ions by electrophysiological techniques that are well known in the art. Untransformed cells, and/or cells transformed with either vector sequences alone or β-galactosidase sequences alone, are used as controls and tested in parallel. The contribution of NTRAN to cation or anion conductance can be shown by incubating the cells using antibodies specific for either NTRAN. The respective antibodies will bind to the extracellular side of NTRAN, thereby blocking the pore in the ion channel, and the associated conductance. To study the dependence of NAP on external ions, sodium can be replaced by choline or N-methyl-D-glucamine and chloride by gluconate, NO_3, or SO_4 (Kavanaugh, M.P. et al. (1992) J. Biol. Chem. 267:22007-22009).

In a further alternative, NTRAN transport activity is assayed by measuring uptake of labeled substrates into *Xenopus laevis* oocytes. Oocytes at stages V and VI are injected with NTRAN mRNA (10 ng per oocyte) and incubated for 3 days at 18 °C in OR2 medium (82.5 mM NaCl, 2.5 mM KCl, 1 mM $CaCl_2$, 1 mM $MgCl_2$, 1 mM Na_2HPO_4, 5 mM Hepes, 3.8 mM NaOH , 50 μg/ml

gentamycin, pH 7.8) to allow expression of NTRAN protein. Oocytes are then transferred to standard uptake medium (100 mM NaCl, 2 mM KCl, 1 mM $CaCl_2$, 1 mM $MgCl_2$, 10 mM Hepes/Tris pH 7.5). Uptake of various substrates (e.g., amino acids, sugars, drugs, and neurotransmitters) is initiated by adding a ^3H substrate to the oocytes. After incubating for 30 minutes, uptake is terminated by washing the oocytes three times in Na^+-free medium, measuring the incorporated ^3H, and comparing with controls. NTRAN activity is proportional to the level of internalized ^3H substrate.

In a further alternative, NTRAN activity can be demonstrated using an electrophysiological assay for ion conductance. Capped NTRAN mRNA transcribed with T7 polymerase is injected into defolliculated stage V *Xenopus* oocytes, similar to the previously described method. Two to seven days later, transport is measured by two-electrode voltage clamp recording. Two-electrode voltage clamp recordings are performed at a holding potential of 50 mV. The data are filtered at 10 Hz and recorded with the MacLab digital-to-analog converter and software for data acquisition and analysis (AD Instruments, Castle Hill, Australia). To study the dependence of NTRAN on external ions, sodium can be replaced by choline or N-methyl-D-glucamine and chloride by gluconate, NO_3, or SO_4 (Kavanaugh, M.P. et al. (1992) J. Biol. Chem. 267:22007-22009).

Calmodulin-binding activity of NTRAN can be demonstrated by mixing 1.5 ml of 0.2 μM calmodulin with increasing concentrations of NTRAN (30 μl in concentrations ranging from 0.5 to 3.1 μM). Analysis of binding activity is determined by Scatchard representation. The bound NTRAN is calculated as $\Delta N/\Delta N_{max200}$ x [total calmodulin]. The free NTRAN concentration is calculated as [total NTRAN] - [bound NTRAN]. (Bosc, C. et al. (2001) J. Biol. Chem. 276:30904-30913).

Alternatively, calmodulin-binding activity of NTRAN can be determined in a binding assay using calmodulin Sepharose beads. Calmodulin Sepharose 6B beads (1 mg/ml calmodulin) are equilibrated in 20 mM Hepes-KOH, pH 7.2, 0.15 M NaCl, 5 mg/ml bovine serum albumin, 0.02% Triton X-100 (v/v) (buffer C) and either 2.5 mM $CaCl_2$ or 5 mM EGTA. Samples of NTRAN are incubated with 5 μl of calmodulin-Sepharose 6B beads for 3 hours at 22°C in 0.5 ml of buffer C in the presence or absence (5 mM EGTA) of 2.5 mM $CaCl_2$. Subsequently, beads are pelleted by centrifugation at 1000 x *g* and washed five times (1.5 ml each wash) with modified buffer C (Triton X-100 concentration is increased to 0.1% (v/v) and albumin is omitted), by resuspension and centrifugation at 1000 x *g*. Washed beads containing bound ligand are incubated in 0.1 ml of SDS gel loading buffer for 5 minutes at 95°C. Bound NTRAN is visualized by Western immunoblotting. (Rossi, E.A. et al. (1999) J. Biol. Chem. 274:27201-27210).

Sialic acid-binding activity of NTRAN can be demonstrated in a COS cell binding assay.

COS cells are transfected with plasmids encoding full-length NTRAN using DEAE-dextran as described in Simmons D.L. (1993) Cloning cell surface molecules by transient expression in mammalian cells. In Cellular Interactions in Development — A Practical Approach. Edited by Hartley D.A.. Oxford: IRL Press, 93–128. On the third day after transfection, binding assays with erythrocytes are carried out for 1 hour at 37 °C in DMEM culture medium with 0.2 % BSA. Unbound erythrocytes are washed off and the cells are fixed in 0.25 % glutardialdehyde. To assess the effect of antibodies on binding, COS cells are pre-incubated with immunoglobulins at 20 μg/ml for 1 hour prior to the addition of erythrocytes (Kelm, S. et al. (1994) Curr. Biol. 4:965-972).

In the alternative, choline transporter activity or choline-transporter-like CTL1 protein activity of NTRAN is determined by measuring choline uptake by yeast transformed with expression vectors harboring polynucleotides encoding NTRAN. The assay is performed in nitrogen-free medium at 30°C for 10 or 30 min in the presence of 25 nM [^{3}H]choline. The cells are then filtered, and washed. The amount of [^{3}H]choline present in the cells is proportional to the activity of NTRAN in the cells (O'Regan, S. *supra*).

In a further alternative, NTRAN protein kinase (PK) activity is measured by phosphorylation of a protein substrate using gamma-labeled [^{32}P]-ATP and quantitation of the incorporated radioactivity using a gamma radioisotope counter. NTRAN is incubated with the protein substrate, [^{32}P]-ATP, and an appropriate kinase buffer. The ^{32}P incorporated into the product is separated from free [^{32}P]-ATP by electrophoresis and the incorporated ^{32}P is counted. The amount of ^{32}P recovered is proportional to the PK activity of NTRAN in the assay. A determination of the specific amino acid residue phosphorylated is made by phosphoamino acid analysis of the hydrolyzed protein.

An assay for NTRAN activity measures the expression of NTRAN on the cell surface. cDNA encoding NTRAN is transfected into a non-leukocytic cell line. Cell surface proteins are labeled with biotin (de la Fuente, M.A. et al. (1997) Blood 90:2398-2405). Immunoprecipitations are performed using NTRAN-specific antibodies, and immunoprecipitated samples are analyzed using SDS-PAGE and immunoblotting techniques. The ratio of labeled immunoprecipitant to unlabeled immunoprecipitant is proportional to the amount of NTRAN expressed on the cell surface.

Alternatively, an assay for NTRAN activity measures the amount of cell aggregation induced by overexpression of NTRAN. In this assay, cultured cells such as NIH3T3 are transfected with cDNA encoding NTRAN contained within a suitable mammalian expression vector under control of a strong promoter. Cotransfection with cDNA encoding a fluorescent marker protein, such as Green Fluorescent Protein (CLONTECH), is useful for identifying stable transfectants. The amount of cell

agglutination, or clumping, associated with transfected cells is compared with that associated with untransfected cells. The amount of cell agglutination is a direct measure of NTRAN activity.

Alternatively, an assay for NTRAN activity measures the disruption of cytoskeletal filament networks upon overexpression of NTRAN in cultured cell lines (Rezniczek, G. A. et al. (1998) J. Cell Biol. 141:209-225). cDNA encoding NTRAN is subcloned into a mammalian expression vector that drives high levels of cDNA expression. This construct is transfected into cultured cells, such as rat kangaroo PtK2 or rat bladder carcinoma 804G cells. Actin filaments and intermediate filaments such as keratin and vimentin are visualized by immunofluorescence microscopy using antibodies and techniques well known in the art. The configuration and abundance of cyoskeletal filaments can be assessed and quantified using confocal imaging techniques. In particular, the bundling and collapse of cytoskeletal filament networks is indicative of NTRAN activity.

Alternatively, cell adhesion activity in NTRAN is measured in a 96-well plate in which wells are first coated with NTRAN by adding solutions of NTRAN of varying concentrations to the wells. Excess NTRAN is washed off with saline, and the wells incubated with a solution of 1% bovine serum albumin to block non-specific cell binding. Aliquots of a cell suspension of a suitable cell type are then added to the wells and incubated for a period of time at 37 °C. Non-adherent cells are washed off with saline and the cells stained with a suitable cell stain such as Coomassie blue. The intensity of staining is measured using a variable wavelength multi-well plate reader and compared to a standard curve to determine the number of cells adhering to the NTRAN coated plates. The degree of cell staining is proportional to the cell adhesion activity of NTRAN in the sample.

Various modifications and variations of the described compositions, methods, and systems of the invention will be apparent to those skilled in the art without departing from the scope and spirit of the invention. It will be appreciated that the invention provides novel and useful proteins, and their encoding polynucleotides, which can be used in the drug discovery process, as well as methods for using these compositions for the detection, diagnosis, and treatment of diseases and conditions. Although the invention has been described in connection with certain embodiments, it should be understood that the invention as claimed should not be unduly limited to such specific embodiments. Nor should the description of such embodiments be considered exhaustive or limit the invention to the precise forms disclosed. Furthermore, elements from one embodiment can be readily recombined with elements from one or more other embodiments. Such combinations can form a number of embodiments within the scope of the invention. It is intended that the scope of the invention be

defined by the following claims and their equivalents.

Table 1

Incyte Project ID	Polypeptide SEQ ID NO:	Incyte Polypeptide ID	Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID	Incyte Full Length Clones
7506758	1	7506758CD1	23	7506758CB1	
4381669	2	4381669CD1	24	4381669CB1	
7503554	3	7503554CD1	25	7503554CB1	5371744CA2
7506077	4	7506077CD1	26	7506077CB1	1977274CA2
7506765	5	7506765CD1	27	7506765CB1	90039685CA2
7490048	6	7490048CD1	28	7490048CB1	
2133585	7	2133585CD1	29	2133585CB1	2133585CA2
7509063	8	7509063CD1	30	7509063CB1	
5496710	9	5496710CD1	31	5496710CB1	
72150826	10	72150826CD1	32	72150826CB1	
6799476	11	6799476CD1	33	6799476CB1	
7509488	12	7509488CD1	34	7509488CB1	
7510060	13	7510060CD1	35	7510060CB1	90087674CA2
7510226	14	7510226CD1	36	7510226CB1	
7510385	15	7510385CD1	37	7510385CB1	
7511618	16	7511618CD1	38	7511618CB1	
6244135	17	6244135CD1	39	6244135CB1	5059410CA2
7506689	18	7506689CD1	40	7506689CB1	95116857CA2
7510185	19	7510185CD1	41	7510185CB1	90196537CA2
1420867	20	1420867CD1	42	1420867CB1	1420867CA2
7512289	21	7512289CD1	43	7512289CB1	
7512447	22	7512447CD1	44	7512447CB1	

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
1	7506758CD1	g13647037	2.3E-56	[Mus musculus] synaptotagmin-like protein 3-a (Fukuda, M. et al. (2001) Biochem. Biophys. Res. Commun. 283:513-519.)
		418460|Sytl4	7.3E-27	[Mus musculus][Small molecule-binding protein] Synaptotagmin-like 4 (granuphilin), a rabphilin-3-like effector of rab family proteins involved in vesicle transport, binds phospholipids in a Ca(2+)-independent manner, may play a role in exocytosis of dense-core granules in endocrine tissues (Wang, J. et al. (1999) J. Biol. Chem. 274:28542-28548; Matesic, L. E. et al. (2001) Proc. Natl. Acad. Sci. U S A 98:10238-10243.)
		432508|RPH3AL	1.8E-11	[Homo sapiens][Cytoplasmic] Putative zinc finger protein with similarity to the N-terminal region of rabphilin-3A, may regulates exocytosis in endocrine cells, is a candidate gene located in 17p13.3 medulloblastoma tumor suppressor locus (Haynes, L. P. et al. (2001) J. Biol. Chem. 276:9726-9732.)
2	4381669CD1	g10880799	0.0	[Mus musculus] Syne-1B (Apel, E. D. et al. (2000) J. Biol. Chem. 275:31986-31995.)
		628322|Syne1	0.0	[Mus musculus][Nuclear] Synaptic nuclear envelope 1, contains multiple spectrin repeats and localizes to the nuclear envelope, enriched in synaptic nuclei at neuromuscular junctions, may function in nuclear migration or anchoring (Apel, E. D. et al. (2000) supra.)
		622033|SYNE-2	1.4E-276	[Homo sapiens][Structural protein][Cytoskeletal] Protein containing spectrin repeats, has similarity to mouse Syne1 and to Drosophila Klarsicht, which is associated with the nuclei of fly photoreceptor cells (Apel, E. D. et al. (2000) supra.)
3	7503554CD1	g13936383	4.7E-184	[Homo sapiens] neuronal leucine-rich repeat protein-3
		748481|FLJ11129	4.1E-185	[Homo sapiens] Protein containing leucine rich repeats, which mediate protein-protein interactions and a leucine rich repeat N-terminal cysteine rich domain (Fukamachi, K. et al. (2001) Biochem. Biophys. Res. Commun. 287:257-263.)

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
3 cont		419907|Lrrn3	2.4E-173	[Mus musculus] Brain leucine rich repeat protein, a member of the leucine-rich repeat family of proteins involved in protein-protein interactions, possibly functions in nervous system development and maintenance (Taniguchi, H. et al. (1996) Brain Res. Mol. Brain Res. 36:45-52; Ishii, N. et al. (1996) Brain Res. Mol. Brain Res. 40:148-152.)
4	7506077CD1	g4102729	8.0E-182	[Homo sapiens] CLN3 protein
		334702|CLN3	7.0E-183	[Homo sapiens][Chaperones][Lysosome/vacuole; Cytoplasmic; Mitochondrial] Ceroid-lipofuscinosis neuronal 3, plays a role in lysosomal pH homeostasis; mutation of the corresponding gene is detected in patients with juvenile neuronal ceroid lipofuscinosis (Batten disease), a lysosomal storage disease (Golabek, A. A. et al. Mol. Genet. Metab. 70:203-213; Lehtovirta, M. et al. (2001) Hum. Mol. Genet. 10:69-75.)
		318588|Cln3	1.1E-151	[Mus musculus][Lysosome/vacuole; Golgi; Cytoplasmic; Mitochondrial] Ceroid-lipofuscinosis neuronal 3, targeted disruption produces mice with a neuronal storage disorder resembling that seen in Batten disease patients; mutation of human CLN3 is detected in patients with juvenile neuronal ceroid lipofuscinosis (Mitchison, H. M. et al. (1999) The Batten Mouse Model Consortium, Neurobiol. Dis. 6:321-334.)
5	7506765CD1	g14278927	3.0E-86	[Mus musculus] gliacolin
		429678|C1qrf	2.5E-73	[Mus musculus] C1q-related factor, has a collagenous region and a globular domain and has similarity to the C1q signature domain, highly expressed in brain areas that are involved in motor function (Berube, N. G. et al. (1999) Brain Res. Mol. Brain Res. 63:233-240.)
		567880|CRF	8.6E-73	[Homo sapiens] C1q-related factor, has a collagenous region and a globular domain and has similarity to the C1q signature domain, highly expressed in brain areas that are involved in motor function (Berube, N. G. et al. (1999) supra.)

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
6	7490048CD1	g6601555	0.0	[Rattus norvegicus] glutamate receptor interacting protein 2 Dong,H.,et al. (1999) J. Neurosci. 19:6930-6941 Characterization of the glutamate receptor-interacting proteins GRIP1 and GRIP2
		734373|Rn.15680	0.0	[Rattus norvegicus] [Anchor Protein] [Unspecified membrane] Glutamate receptor interacting protein 2, an adaptor protein that has two sets of three PDZ domains, may act as a scaffold protein that anchors AMPA glutamate receptors at excitatory synapses and is localized within lipid rafts of the plasma membrane
		425390|GRIP1	3.20E-237	[Homo sapiens] [Structural protein; DNA-binding protein; Transcription factor] [Nuclear; Plasma membrane] Glutamate receptor interacting protein, transcriptional coactivator involved in estrogen signaling, interacts specifically with ephrin B1 (EPHB1) to form heterodimers that make a scaffolding framework for multiprotein signaling complexes formation Shang, Y., et al. (2000) Cell 103:843-852 Cofactor dynamics and sufficiency in estrogen receptor-regulated transcription.
7	2133585CD1	g1685069	5.2E-226	[Homo sapiens] sushi-repeat-containing protein precursor Meindl, A.et al. (1995) Hum. Mol. Genet. 4:2339-2346 A gene (SRPX) encoding a sushi-repeat-containing protein is deleted in patients with X-linked retinitis pigmentosa.
		343818|SRPX	4.50E-227	[Homo sapiens] [Unspecified membrane] Sushi-repeat-containing protein (X chromosome), may be involved in negative regulation of cell growth, expression is reduced in several cancer cell lines; mutation of the corresponding gene may be associated with X-linked retinitis pigmentosa (XLRP) Shimakage, M. et al. (2000) Int. J. Cancer 87:5-11 Down-regulation of drs mRNA in human colon adenocarcinomas.

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
7 cont		629524|Drs	7.40E-211	[Rattus norvegicus] [Plasma membrane] Sushi-repeat-containing protein (X chromosome), protein suppressed in v-src-transformed cells, may be involved in inhibition of transformation; mutation of the human SRPX may be associated with X-linked retinitis pigmentosa (XLRP)
8	7509063CD1	g10880799	0.0	[Mus musculus] Syne-1B Apel, E.D. (2000) J. Biol. Chem. 275:31986-31995 Syne-1, A Dystrophin- and Klarsicht-related Protein Associated with Synaptic Nuclei at the Neuromuscular Junction.
		628322|Syne1	0.0	[Mus musculus] [Nuclear] Synaptic nuclear envelope 1, contains multiple spectrin repeats and localizes to the nuclear envelope, enriched in synaptic nuclei at neuromuscular junctions, may function in nuclear migration or anchoring
		622033|SYNE-2	1.90E-268	[Homo sapiens] [Structural protein] [Cytoskeletal] Protein containing spectrin repeats, has similarity to mouse Syne1 and to Drosophila Klarsicht, which is associated with the nuclei of fly photoreceptor cells
9	5496710CD1	g18390059	0.0	cell recognition protein CASPR4 [Homo sapiens]
		g12330704	0.0	[Mus musculus] cell recognition molecule CASPR4
		336668|CNTNAP1	4.20E-255	[Homo sapiens] [Receptor (signalling)] [Unspecified membrane; Plasma membrane] Neurexin 4 (contactin associated protein 1), a neuronal paranodal transmembrane receptor containing epidermal growth factor-like and neurexin motifs, binds contactin (CNTN1), may have a role in neuronal intracellular signaling Peles, E. et al. (1997) EMBO J. 16:978-988 Identification of a novel contactin-associated transmembrane receptor with multiple domains implicated in protein-protein interactions.
		587769|Nrxn4	2.30E-252	[Mus musculus] [Receptor (signalling)] [Unspecified membrane; Plasma membrane] Neurexin 4 (contactin associated protein 1), a putative neuronal paranodal transmembrane receptor, may have a role in neuronal intracellular signaling; mutant animals display defects in axonogenesis

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
10	72150826CD1	g211123	2.9E-94	[Gallus gallus] agrin-related protein 1 Ruegg, M.A.(1992) Neuron 8:691-699 The agrin gene codes for a family of basal lamina proteins that differ in function and distribution.
		328066|Agrn	5.50E-90	[Rattus norvegicus] [Inhibitor or repressor] [Extracellular matrix (cuticle and basement membrane); Basement membrane (extracellular matrix); Extracellular (excluding cell wall)] Agrin, proteoglycan component of the basal lamina that stimulates acetylcholine receptor aggregation through the activation of MUSK, involved in synaptogenesis
		310445|AGRN	4.30E-89	[Homo sapiens] [Inhibitor or repressor] [Extracellular matrix (cuticle and basement membrane); Basement membrane (extracellular matrix); Extracellular (excluding cell wall)] Agrin, a component of the basement membrane, mediates acetylcholine receptor clustering, likely through the activation of MUSK, may promote amyloidogenic peptide aggregation in Alzheimer's disease Berzin, T. M. et al. (2000) Neurobiol. Aging 21:349-355 Agrin and microvascular damage in Alzheimer's disease.
11	6799476CD1	g11558246	0.0	[Mus musculus] calsyntenin-1 protein Vogt, L., et al. (2001) Mol. Cell. Neurosci. 17:151-166 Calsyntenin-1, a Proteolytically Processed Postsynaptic Membrane Protein with a Cytoplasmic Calcium-Binding Domain.
		570390|KIAA0911	0.0	[Homo sapiens] Protein containing cadherin domains, which are found in the extracellular domain of some receptors and cell adhesion proteins
		626171|CS2	6.50E-299	[Homo sapiens] Protein containing two cadherin domains, which are found in the extracellular domain of some receptors and cell adhesion proteins, has strong similarity to uncharacterized calsyntenin-2 (mouse Cstn2)
12	7509488CD1	g11558246	0.0	[Mus musculus] calsyntenin-1 protein Vogt,L., et al., supra

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
12 cont		570390|KIAA0911	0	[Homo sapiens] Protein containing cadherin domains, which are found in the extracellular domain of some receptors and cell adhesion proteins
		626171|CS2	8.90E-302	[Homo sapiens] Protein containing two cadherin domains, which are found in the extracellular domain of some receptors and cell adhesion proteins, has strong similarity to uncharacterized calsyntenin-2 (mouse Cstn2)
13	7510060CD1	g307156	0.0	[Homo sapiens] myelin-associated glycoprotein precursor Sato, S. et al. (1989) Biochem. Biophys. Res. Commun. 163:1473-1480 cDNA cloning and amino acid sequence for human myelin-associated glycoprotein.
		623760|MAG	0.0	[Homo sapiens][Plasma membrane] Myelin associated glycoprotein, a member of the sialoadhesin subgroup of immunoglobulin superfamily lectins which may play a role in cell adhesion; identified as an autoantigen in several forms of demyelinating neuropathy including multiple sclerosis Miescher, G. C. et al. (1997) Brain Res Mol Brain Res 52:299-306 Reciprocal expression of myelin-associated glycoprotein splice variants in the adult human peripheral and central nervous systems.
		591203|Mag	0.0	[Rattus norvegicus][Plasma membrane] Myelin-associated glycoprotein, a lectin which inhibits axon outgrowth and may play a role in cell adhesion; identified as an autoantigen in several forms of demyelinating neuropathy Royland, J. et al. (1992) Brain Res Dev Brain Res 65: 223-226 Down regulation of myelin-specific mRNAs in the mechanism of hypomyelination in the undernourished developing brain.
14	7510226CD1	g6851362	4.3E-93	[Homo sapiens] zinedin Castets, F. et al. (2000) J. Biol. Chem. 275, 19970-19977 Zinedin, SG2NA, and striatin are calmodulin-binding, WD repeat proteins principally expressed in the brain.

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
14 cont		432836|ZIN	3.5E-94	[Homo sapiens][Structural protein;Small molecule-binding protein][Cytoplasmic] Zinedin, contains WD-repeats and has similarity to the putative scaffolding protein striatin (STRN), binds calmodulin (CALM1) in a calcium sensitive-manner Castets, F. et al. supra
		609919|Strn	6.4E-45	[Rattus norvegicus][Small molecule-binding protein] Striatin, binds calmodulin in the presence of calcium, contains WD-repeats, possibly functions in excitatory synapse calcium-signaling pathways Castets, F. et al. (supra)
15	7510385CD1	g974284	7.3E-169	[Homo sapiens] semaphorin V Wei, M. H. et al. (1996) Cancer Res. 56, 1487-1492 Construction of a 600-kilobase cosmid clone contig and generation of a transcriptional map surrounding the lung cancer tumor suppressor gene (TSG) locus on human chromosome 3p21.3: progress toward the isolation of a lung cancer TSG
16	7511618CD1	g18254511	4.6E-185	[Homo sapiens] ubiquilin
				Mah, A. L. et al.
				Identification of ubiquilin, a novel presenilin interactor that increases presenilin protein accumulation
				J. Cell Biol. 151, 847-862 (2000)
		437284|UBQLN1	5.5E-185	[Homo sapiens][Endoplasmic reticulum;Nuclear;Cytoplasmic] Ubiquilin 1, a ubiquitin-related protein associated with the 26S proteasome, binds to and promotes the accumulation of PSEN1 and PSEN2, localizes to neurofibrillary tangles and Lewy bodies in brains affected by Alzheimer's disease and Parkinson's disease
				Kleijnen, M. F. et al.
				The hPLIC proteins may provide a link between the ubiquitination machinery and the proteasome.
				Mol Cell 6, 409-19 (2000).

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
16 cont		757282|Ubqln1	1.3E-160	[Rattus norvegicus] Protein with very strong similarity to ubiquilin 1 (mouse Ubqln1), which regulates the interaction between the plasma membrane and vimentin intermediate filaments, member of the ubiquitin family, contains two UBA (ubiquitin associated) or TS-N domains
				Bedford, F. K. et al. Nat. Neurosci. 4, 908-916 (2001).
17	6244135CD1	g200250	6.9E-40	[Mus musculus] PCD-5
				Nordquist, D. T. et al.
				cDNA cloning and characterization of three genes uniquely expressed in cerebellum by Purkinje neurons
				J. Neurosci. 8, 4780-4789 (1988)
		582377|Pcp2	5.6E-41	[Mus musculus] Purkinje cell protein 2, a novel guanyl nucleotide exchange factor that interacts with G protein Gnao, may have a role in retinal and central nervous system development, expression is confined to Purkinje cells within the cerebellum
				Sandhofer, C. et al.
				Beta receptor isoforms are not essential for thyroid hormone-dependent acceleration of PCP-2 and myelin basic protein gene expression in the developing brains of neonatal mice.
				Mol Cell Endocrinol 137, 109-15 (1998).
18	7506689CD1	g23468354	6.0E-65	sema domain, immunoglobulin domain (Ig), transmembrane domain (TM) and short cytoplasmic domain, (semaphorin) 4F [Homo sapiens]
		g4519503	3.0E-76	[Homo sapiens] semaphorin W
				Encinas, J. A. et al.
				Cloning, expression, and genetic mapping of Sema W, a member of the semaphorin family
				Proc. Natl. Acad. Sci. U.S.A. 96, 2491-2496 (1999)
		341226|SEMA4F	2.5E-77	[Homo sapiens][Plasma membrane] Semaphorin 4F (semaphorin W), a member of the semaphorin family, may be involved in neuronal growth cone guidance

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
18 cont				Encinas, J. A. et al. supra
		609873|Sema4f	1.7E-61	[Rattus norvegicus][Plasma membrane] Semaphorin 4F(semaphorin W), a member of the semaphorin family, involved in neuronal growth cone guidance, induces the collapse of growth cones
				Schultze, W. et al.
				Semaphorin4F interacts with the synapse-associated protein SAP90/PSD-95.
				J Neurochem 78, 482-9. (2001).
19	7510185CD1	g3068592	0.0	[Mus musculus] punc Salbaum, J. M. Mech. Dev. 71, 201-204 (1998) Salbaum, J. M. Mamm. Genome 10, 107-111 (1999)
		582679|Punc	0.0	[Mus musculus][Adhesin/agglutinin][Plasma membrane] Putative neuronal cell adhesion molecule, a member of the immunoglobulin superfamily of cell surface proteins, may play a role in cerebellar control of motor coordination and early embryogenesis Salbaum, J. M. Mech Dev 71, 201-4 (1998). Yang, W. et al. Mol Cell Biol 21, 6031-43 (2001).
		348156|PUNC	2.0E-162	[Homo sapiens][Adhesin/agglutinin][Plasma membrane] Putative neuronal cell adhesion molecule, a member of the immunoglobulin superfamily of cell surface proteins, highly expressed in the embryonic nervous system and limb buds Salbaum, J. M (1999, supra)
20	1420867CD1	g13235326	1.8E-89	[Homo sapiens] dJ477O4.1.2 (novel protein similar to otoferlin and dysferlin, isoform 2)

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
20 cont		341002|OTOF	1.3E-38	[Homo sapiens][Cytoplasmic; Unspecified membrane] Otoferlin, may play a role in synaptic or other vesicle membrane fusion and is required for inner ear function; alteration of the corresponding gene is associated with nonsyndromic prelingual deafness Yasunaga, S. et al. Nat Genet 21, 363-9 (1999). Yasunaga, S. et al. Am J Hum Genet 67, 591-600 (2000).
		710087|Otof	5.7E-38	[Mus musculus] Otoferlin, may play a role in synaptic or other vesicle membrane fusion; alteration of human OTOF is associated with nonsyndromic prelingual deafness Yasunaga, S. et al (1999, supra)
21	7512289CD1	g7658295	5.9E-122	[Homo sapiens] transmembrane protein BRI Vidal, R. et al. Proc. Natl. Acad. Sci. U.S.A. 97, 4920-4925 (2000)
		626041|ITM2B	1.8E-123	[Homo sapiens][Unspecified membrane; Plasma membrane] Integral membrane protein 2B, a type II integral membrane protein; gene mutation is associated with familial British and Danish dementias, which are characterized by amyloid plaque deposits, neurofibrillary tangles, and neuronal loss El-Agnaf, O. M. et al. Biochemistry 40, 3449-57 (2001). Holton, J. L. et al Am J Pathol 158, 515-26 (2001).
		583443|Itm2b	1.2E-77	[Mus musculus][Unspecified membrane] Integral membrane protein 2B, a type II integral membrane protein; human ITM2B mutation is associated with familial British and Danish dementias, which are characterized by amyloid plaque deposits, neurofibrillary tangles, and neuronal loss Pittois, K. et al Gene 217, 141-9 (1998).
22	7512447CD1	g14149046	2.6E-67	[Drosophila melanogaster] turtle protein, isoform 2 Bodily, K. D. et al. Neuroscience. 21, 3113-3125 (2001)

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
22 cont		590029|Dcc	2.4E-28	[Rattus norvegicus][Receptor (signalling)][Plasma membrane] Deleted in colorectal cancer, netrin 1 receptor, member of the immunoglobulin-CAM family which mediates axon guidance and regulates apoptosis; human DCC is deleted in colorectal cancers Zou, Y. et al. Cell 102, 363-75 (2000). Stein, E. et al. Science 291, 1976-82 (2001).
		742836|DSCAM	3.6E-20	[Homo sapiens][Adhesin/agglutinin][Plasma membrane] Down syndrome cell adhesion molecule, a member of the immunoglobulin superfamily, expressed primarily in the brain Agarwala, K. L. et al. Brain Res Mol Brain Res 79, 118-26 (2000). Yamakawa, K. et al. Hum Mol Genet 7, 227-37 (1998).

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
1	7506758CD1	151	Phorbol esters / diacylglycerol binding domain: C63-K127	PROFILESCAN
			Potential Phosphorylation Sites: S84 T31 T36 T54 Y121	MOTIFS
2	4381669CD1	2507	Spectrin repeat: Q2150-D2257, L1062-S1163, G1819-R1923, N1491-E1593, E732-S837, K505-Q614, V1596-E1707, S1408-E1488, S872-Q940, L1166-S1271, R1710-H1816, E704-G729, T2120-Q2147	HMMER_PFAM
			Spectrin repeat proteins PF00435: F1403-A1431, W1577-K1592, L1993-A2012 $P < 0.0052$	BLIMPS_PFAM
			Potential Phosphorylation Sites: S63 S86 S176 S209 S266 S286 S299 S328 S329 S402 S448 S471 S473 S523 S564 S592 S634 S688 S694 S697 S733 S735 S837 S844 S943 S1348 S1369 S1408 S1410 S1434 S1454 S1480 S1497 S1564 S1607 S1720 S1729 S1788 S1849 S1933 S1977 S1980 S2015 S2039 S2079 S2092 S2096 S2246 S2291 S2321 S2375 S2383 S2384 S2398 S2429 S2437 S2446 S2478 T102 T296 T314 T490 T586 T721 T746 T813 T922 T980 T1005 T1098 T1155 T1170 T1177 T1313 T1402 T1508 T1673 T1708 T1765 T1808 T1826 T1836 T1843 T2072 T2169 T2220 T2408 Y396 Y1128 Y1295 Y1568	MOTIFS
			Potential Glycosylation Sites: N294 N681 N710 N1604 N1781 N1966	MOTIFS

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
			Cell attachment sequence: R256-D258	MOTIFS
3	7503554CD1	375	signal_cleavage: M1-A22	SPSCAN
			Signal Peptide: M4-A22, M1-A22	HMMER
			Leucine Rich Repeat: N261-L284, N213-E236, N93-P116, D310-P334, N237-V260, N189-I212, H285-P309, N141-H164, N165-P188, Q117-S140	HMMER_PFAM
			Leucine rich repeat N-terminal domain: D28-P68	HMMER_PFAM
			Leucine-rich repeat signature PR00019: L94-V107, L139-L152	BLIMPS_PRINTS
			PROTEIN BAC CLONE RG118D07 FROM 7Q31 COMPLETE SEQUENCE GLIOMA AMPLIFIED PD032016: M283-L349	BLAST_PRODOM
			PROTEIN BAC CLONE RG118D07 FROM 7Q31 COMPLETE SEQUENCE NEURONAL LEUCINERICH PD080235: P188-E236	BLAST_PRODOM
			BAC CLONE RG118D07 FROM 7Q31 COMPLETE SEQUENCE GLIOMA AMPLIFIED ON PD080342: N237-N282	BLAST_PRODOM
			Potential Phosphorylation Sites: S51 S173 S242 T42 T95	MOTIFS
			Potential Glycosylation Sites: N93 N103 N223	MOTIFS

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
			Cell attachment sequence: R277-D279	MOTIFS
4	7506077CD1	340	CLN3 protein: L99-S340, M1-A98	HMMER_PFAM
			Cytosolic domains: M1-N37, T134-G182, R240-R245, H296-R307 Transmembrane domains: A38-H60, T114-L133, L183-F205, Q220-L239, F246-G268, I273-F295, E308-L330 Non-cytosolic domains: D61-Q113, E206-Q219, F269-S272, P331-S340	TMHMMER
			PROTEIN TRANSMEMBRANE CLN3 BTN1 BTN1P BATTEN'S DISEASE NEURONAL CEROID LIPOFUSCINOSIS PD007801: D140-L339, D13-S135, A98-K164	BLAST_PRODOM
			CLN3 PROTEIN TRANSMEMBRANE BATTEN'S DISEASE NEURONAL CEROID LIPOFUSCINOSIS MUTATION PD156246: M1-K36	BLAST_PRODOM
			CLN3; DISEASE; BATTEN; DM05338|A57219|83-438: A98-S340, P83-L124 DM05338|P47040|1-407: Y186-Q338, F41-L63	BLAST_DOMO
			Potential Phosphorylation Sites: S7 S12 S14 S64 S69 S73 S87 S88 S172 S238 T19 T134 T302	MOTIFS

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
			Potential Glycosylation Sites: N49 N71 N85 N212	MOTIFS
5	7506765CD1	287	signal_cleavage: M1-A15	SPSCAN
			Signal Peptide: M1-A20, M1-A21, M1-H22	HMMER
			C1q domain: A160-L284	HMMER_PFAM
			Collagen triple helix repeat (20 copies): G76-G135	HMMER_PFAM
			C1q domain proteins BL01113: G85-S111, V174-I209, D243-K262, S277-P286	BLIMPS_BLOCKS
			Lysosome-associated membrane glycoproteins signatures: E145-L175	PROFILESCAN
			Complement C1Q domain signature PR00007: P168-K194, F195-G214, D243-D264, K275-Y285	BLIMPS_PRINTS
			PRECURSOR SIGNAL COLLAGEN REPEAT HYDROXYLATION GLYCOPROTEIN CHAIN PLASMA EXTRACELLULAR MATRIX PD002992: P168-L284	BLAST_PRODOM
			COLLAGEN ALPHA PRECURSOR CHAIN REPEAT SIGNAL CONNECTIVE TISSUE EXTRACELLULAR MATRIX PD000007: A39-G118, P72-G118, P35-D110	BLAST_PRODOM
			PRECOLLAGEN P PRECURSOR SIGNAL PD072959: A38-G118, P40-P117	BLAST_PRODOM
			PROCOLLAGEN, TYPE XVII, ALPHA 1 BULLOUS PEMPHIGOID AUTOANTIGEN BP180 COLLAGEN XVII ANTIGEN CELL ADHESION PD071338: P44-G112	BLAST_PRODOM
			C1Q DOMAIN :DM00777|S23297|465-674: P84-L283, P69-G213; DM00777|P23206|477-673: G94-P286; DM00777|P02746|70-250: G82-D287;DM00777|P27658|551-743: G88-P286	BLAST_DOMO
			Potential Phosphorylation Sites: S111 S142 S167 S255 T28 T192	MOTIFS

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
			C1q domain signature: F177-Y207	MOTIFS
6	7490048CD1	1048	PDZ domain (Also known as DHR or GLGF): T661-Q742, H562-D645, D153-D238, V53-E135, E461-D549, M253-V336, K946-K1027	HMMER_PFAM
			PDZ domain proteins (Also known as DHR or GLGF) PF00595: I702-N712	BLIMPS_PFAM
			PROTEIN SH3 DOMAIN REPEAT PD00289: G705-G718	BLIMPS_PRODOM
			AMPA Receptor Protein PD084959: P335-T460; PD065834: A845-P1040; PD176166: I279-L334; *PD117239: L748-Y840*	BLAST_PRODOM
			Potential Phosphorylation Sites: S5 S43 S197 S272 S288 S314 S324 S326 S339 S369 S422 S499 S570 S581 S597 S676 S715 S754 S759 S787 S795 S839 S848 S929 S961 S1025 S1033 S1044 T52 T72 T223 T308 T525 T691 T736 T817 T843 T940 T975 Y23 Y134 Y157	MOTIFS
			Potential Glycosylation Sites: N247 N275 N713	MOTIFS
7	2133585CD1	419	Signal_cleavage: M1-S30	SPSCAN
			Signal Peptide: H6-S30, M1-S30, L13-S30, P8-S30	HMMER
			HYR (hyalin repeat) domain: V175-V257	HMMER_PFAM
			Sushi domain (SCR repeat): C57-C115, C262-C317, C120-C174	HMMER_PFAM
			Sushi domain proteins (S PF00084): C69-Q73, N140-Y151, L308-C317	BLIMPS_PFAM
			PROTEIN SUSHI REPEATCONTAINING SRPX PRECURSOR SIGNAL REPEAT SUSHI REPEAT POLYMORPHISM ALTERNATIVE PD020056: V25-S311; PD019555: V175-R261; PD021630: A318-L403	BLAST_PRODOM

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
			SUSHI REPEAT DM04887 P33730|1-610: G65-E189, K63-M320, G65-C174, Y282-A319, R261-C317; P16581|1-609: G45-M320, L14-Y51 P27113|1-551: G45-C317, C57-C174, G65-V187, R258-V326, G381-L416	BLAST_DOMO
			Potential Phosphorylation Sites: S37 S105 S110 S141 S168 S173 S186 S277 S311 T83 T152 T160 T198 T240 T250 T349 T375 T409 Y138 Y282	MOTIFS
8	7509063CD1	1109	Spectrin repeat: T722-Q749, Q752-D859, G397-R501, N69-E171, V174-E285, R2-E66, R288-H394	HMMER_PFAM
			Potential Phosphorylation Sites: S12 S32 S58 S75 S142 S185 S298 S307 S366 S427 S512 S556 S559 S594 S641 S681 S694 S698 S848 S893 S923 S977 S985 S986 S1000 S1031 S1039 S1048 S1080 T86 T251 T286 T343 T386 T404 T414 T421 T636 T674 T771 T822 T1010 Y146	MOTIFS
			Potential Glycosylation Sites: N182 N359 N545 N635	MOTIFS
9	5496710CD1	1308	Signal Peptide: M1-A25	HMMER
			EGF-like domain: C553-C585, C962-C996, R552-H586, H961-S997	HMMER_PFAM
			F5/8 type C (coagulation-associated) domain: P34-V174, D30-C177	HMMER_PFAM
			Laminin G domain: L813-M940, I204-V341, E390-K524, S1038-V1176, F821-D943, F398-D527, F212-I344, N1073-S1131	HMMER_PFAM
			Cytosolic domain: Y1265-F1308 Transmembrane domain: V1242-I1264 Non-cytosolic domain: M1-A1241	TMHMMER

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
			PROTEIN NEUREXIN IV CONTACTIN ASSOCIATED PD013624: C558-V817; PD022248: S818-S1030; PD022247: V341-S550	BLAST_PRODOM
			W03D8.6 PROTEIN GLYCOPROTEIN PD185819: F928-S1169, C785-D983, G511-G601, S396-S535, I208-G314	BLAST_PRODOM
			DISCOIDIN I N-TERMINAL DM00516|A42580|2085-2210: P56-C177; DM00516|P12259|2095-2223: P56-C177 DM00516|P21956|338-462: P56-C177; DM00516|P00451|2221-2347: P56-C177	BLAST_DOMO
			Potential Phosphorylation Sites: S46 S75 S149 S209 S265 S376 S387 S459 S522 S550 S588 S715 S1045 S1131 S1150 S1186 S1218 T287 T647 T769 T896 T1048 T1211 T1225 Y711 Y778	MOTIFS
			Potential Glycosylation Sites: N260 N285 N359 N538 N574 N602 N625 N637 N706 N748 N1023 N1073	MOTIFS
			Cell attachment sequence: R786-D788	MOTIFS
			ATP/GTP-binding site motif A (P-loop): G314-S321	MOTIFS
			Coagulation factors 5/8 type C domain (FA58C) signature 1: G69-G98	MOTIFS
			Coagulation factors 5/8 type C domain (FA58C) signature 2: P159-C177	MOTIFS
			Leucine zipper pattern: L407-L428	MOTIFS
10	72150826CD1	1009	Signal_cleavage: M1-S24	SPSCAN
			Signal Peptide: M1-S17, M1-G19, M1-G21, M1-A22, M1-S24, M1-R26	HMMER
			EGF-like domain: C569-C601, C788-C819, C347-C380	HMMER_PFAM

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
			Fibronectin type III domain: P142-S228, V35-S126	HMMER_PFAM
			Laminin G domain: F415-D546, F641-H770, F860-I980	HMMER_PFAM
			Cytosolic domain: M1-G6 Transmembrane domain: V7-I29 Non-cytosolic domain: R30-K1009	TMHMMER
			Type II EGF-like signature PR00010: D560-E571, G598-F605, D586-F596, K597-D603	BLIMPS_PRINTS
			NEUREXIN; ALPHA; III; CYSTEINE; DM00060|Q05793|3542-3652: K861-G960, M635-G738 DM00060|P98160|4227-4338: M858-G960, E638-G738, I411-A515 DM00060|Q05793|3004-3118: Y631-G738, Y406-G514 DM00060|Q05793|3265-3373: M635-V739, Y406-G514, M858-M961	BLAST_DOMO
			Potential Phosphorylation Sites: S24 S80 S127 S154 S161 S191 S225 S244 S273 S294 S298 S379 S557 S615 S633 S749 S798 S850 S911 S928 S992 T96 T190 T505 T527 T640 T703 T863 T932 T940 T968	MOTIFS
			Potential Glycosylation Sites: N47 N909	MOTIFS
			Cell attachment sequence: R434-D436, R872-D874	MOTIFS
			EGF-like domain signature 1: C369-C380, C590-C601, C808-C819	MOTIFS
			EGF-like domain signature 2: C590-C601, C808-C819	MOTIFS
11	6799476CD1	962	Signal_cleavage: M1-A28	SPSCAN
			Signal Peptide: A10-A28, L9-A28, P7-A28, M1-G24, M1-A28, P11-A28, P5-A28, P7-A28	HMMER
			Cadherin domain: Y169-K258, Y42-N155	HMMER_PFAM
			Cytosolic domain: R864-Y962 Transmembrane domain: A841-I863 Non-cytosolic domain: M1-T840	TMHMMER

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
			Potential Phosphorylation Sites: S104 S145 S168 S256 S320 S381 S408 S483 S493 S527 S554 S649 S732 S748 S765 S819 S904 S905 S934 T47 T172 T222 T238 T248 T291 T318 T403 T477 T524 T541 T607 T684 T741 T870 T876 T898 T927 Y42 Y227 Y319 Y540	MOTIFS
			Potential Glycosylation Sites: N346 N366 N946	MOTIFS
12	7509488CD1	953	Signal_cleavage: M1-A29	SPSCAN
			Signal Peptide: P11-A29, V10-A29, R8-A29, P5-A29, M1-A29, P12-A29	HMMER
			Cadherin domain: Y160-K249, Y43-N146	HMMER_PFAM
			Cytosolic domain: R855-Y953 Transmembrane domain: A832-I854 Non-cytosolic domain: M1-T831	TMHMMER
			Potential Phosphorylation Sites: S95 S136 S159 S247 S311 S372 S399 S474 S484 S518 S545 S640 S723 S739 S756 S810 S895 S896 S925 T48 T163 T213 T229 T239 T282 T309 T394 T468 T515 T532 T598 T675 T732 T861 T867 T889 T918 Y43 Y218 Y310 Y531	MOTIFS
			Potential Glycosylation Sites: N337 N357 N937	MOTIFS
13	7510060CD1	582	Signal Peptide: M1-G23, M1-G19	HMMER
			I type Ig domains from SCOP: P239-P329, V331-L416	HMMER_INCY
			Ig superfamily from SCOP: K328-L415, I242-L320, T140-P239	HMMER_INCY
			Immunoglobulin: N332-E410, N246-M323, P27-L135, V144-K237	HMMER_SMART
			Immunoglobulin C-2 Type: I252-G312, V338-G399	HMMER_SMART
			Immunoglobulin domain: G340-A394, G254-A307, G35-Y116	HMMER_PFAM

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
			Cytosolic domain: T534-H582 Transmembrane domain: W511-I533 Non-cytosolic domain: M1-M510	TMHMMER
			GLYCOPROTEIN ANTIGEN PRECURSOR PD02327: L259-L280	BLIMPS_PRODOM
			GLYCOPROTEIN PRECURSOR MYELIN CELL ADHESION TRANSMEMBRANE SIGNAL IMMUNOGLOBULIN FOLD MYELINASSOCIATED PD017076: E395-E580	BLAST_PRODOM
			CELL PRECURSOR GLYCOPROTEIN TRANSMEMBRANE SIGNAL IMMUNOGLOBULIN FOLD ADHESION ALTERNATIVE SPLICING PD005007: L7-E229	BLAST_PRODOM
			do MYELIN; SCHWANN; SIALOADHESIN; FORM; DM03744|P20917|1-138: M1-N139 DM03744|JH0593|1-137: M1-I137	BLAST_DOMO
			IMMUNOGLOBULIN DM00001|P20917|140-234: T140-D235 DM00001|P20917|323-407: Y238-S408	BLAST_DOMO
			Potential Phosphorylation Sites: S31 S133 S247 S351 S383 S436 S456 S463 S549 S577 T128 T153 T296 T452 T475 T489 T536	MOTIFS
			Potential Glycosylation Sites: N99 N223 N246 N315 N332 N406 N450 N454 N541	MOTIFS
			Cell attachment sequence: R118-D120	MOTIFS
14	7510226CD1	188	signal_cleavage: M1-C17	SPSCAN
			Signal Peptide: M2-A24	HMMER
			ZP domain proteins BL00682: S15-G21	BLIMPS_BLOCKS
			Ribosomal protein S15 signature: Q73-G144	PROFILESCAN
			Potential Phosphorylation Sites: S16	MOTIFS
			REPEAT PROTEIN STRIATIN WD CALMODULINBINDING CELLCYCLE NUCLEAR AUTOANTIGEN SG2NA S/G2 PD011690: T57-R179	BLAST_PRODOM

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
15	7510385CD1	478	signal_cleavage: M1-A24	SPSCAN
			Signal Peptide: M1-G22, M1-A24, M1-S27	HMMER
			Sema domain: Y55-R308	HMMER_PFAM
			Cytosolic domain: M1-A4 Transmembrane domain: G5-S27 Non-cytosolic domain: P28-R478	TMHMMER
			SEMAPHORIN PROTEIN PRECURSOR RECEPTOR KINASE SIGNAL TYROSINE TYROSINEPROTEIN FAMILY HEPATOCYTE PD001844: Y55-E183, E162-G311	BLAST_PRODOM
			SEMAPHORIN V A PRECURSOR SEM SIGNAL IMMUNOGLOBULIN FOLD MULTIGENE FAMILY PD022826: M1-C54	BLAST_PRODOM
			do SEMAPHORIN; FASCICLIN; COLLAPSIN; II; DM01606|I48744|1-639: M1-E309 DM01606|I48747|1-646: G5-E309 DM01606|A49069|1-646: G5-E309 DM01606|I48748|1-589: A6-E309	BLAST_DOMO
			Potential Phosphorylation Sites: S34 S84 S168 S209 S321 S327 T249 T302	MOTIFS
			Potential Glycosylation Sites: N82 N124	MOTIFS
16	7511618CD1	412	UBA/TS-N domain: R369-S409	HMMER_PFAM
			Ubiquitin associated domain: F370-G408	HMMER_SMART
			F15C11.2 PROTEIN, Ubiquitin family (*C. elegans*) PD183489: V58-P330	BLAST_PRODOM
			Potential Phosphorylation Sites: S83 S127 S135 T40 T43 T329 T395	MOTIFS
			Potential Glycosylation Sites: N55 N125 N324	MOTIFS
17	6244135CD1	136	LGN motif, putative GEF specific for G-alpha: M63-V85, Q23-L45	HMMER_PFAM

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
			LGN motif, putative GEFs specific for G-alpha: M63-V85, Q23-L45	HMMER_SMART
			signal_cleavage: M38-S96	SPSCAN
			PURKINJE CELL PROTEIN 2 PCD5 CELL SPECIFIC L7 NEURONE PD067101: M38-P136	BLAST_PRODOM
			Potential Phosphorylation Sites: S55 S65 S127 T52 T114	MOTIFS
18	7506689CD1	198	signal_cleavage: M1-G34	SPSCAN
			Signal Peptide: M1-G34	HMMER
			Potential Phosphorylation Sites: S33 S48 S153 S178	MOTIFS
			Potential Glycosylation Sites: N64	MOTIFS
19	7510185CD1	813	signal_cleavage: M1-G35	SPSCAN
			Signal Peptide: P18-G35	HMMER
			Signal Peptide: L16-G35	HMMER
			Signal Peptide: P13-G35	HMMER
			Signal Peptide: R11-G35	HMMER
			Signal Peptide: M1-G35	HMMER
			Signal Peptide: A14-G35	HMMER
			Signal Peptide: P13-G37	HMMER
			Fibronectin type 3 domain: P424-S507, A521-G603	HMMER_SMART
			Immunoglobulin: P336-L418, P244-Q327, P145-S228, P48-A136	HMMER_SMART
			Immunoglobulin C-2 Type: P342-G407, V54-G124, E151-S216, T250-T314	HMMER_SMART
			Fibronectin type III domain: P424-S510, A521-T606	HMMER_PFAM
			Immunoglobulin domain: G344-A402, H252-A309, G56-A119, G153-A211	HMMER_PFAM
			I type Ig domains from SCOP: H335-G425, K236-F332, E41-D140	HMMER_INCY
			Ig superfamily from SCOP: F332-L423, F141-G231, I240-A330, F44-I133	HMMER_INCY

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
			Cytosolic domain: G662-Q813 Transmembrane domain: T639-F661 Non-cytosolic domain: M1-T638	TMHMMER
			NEURONAL CELL ADHESION MOLECULE PUNC PD123980: N604-L621 T639-A803	BLAST_PRODOM
			BASIC FIBROBLAST GROWTH FACTOR RECEPTOR 1 DM01287|A39752|1-814: L16-V118 P244-E421 DM01287|P16092|1-821: W17-G124 P244-A420 DM01287|A39752|1-814: Q134-V210	BLAST_DOMO
			IMMUNOGLOBULIN DM00001|P43146|328-410: P329-Q410	BLAST_DOMO
			Potential Phosphorylation Sites: S8 S108 S128 S216 S272 S412 S436 S443 S486 S528 S582 S611 S616 S685 T74 T137 T148 T181 T502 T517 T606	MOTIFS
			Potential Glycosylation Sites: N93 N246 N381 N382 N580 N604 N634	MOTIFS
20	1420867CD1	165	FER1 PD185307: M1-Q125	BLAST_PRODOM
			Potential Phosphorylation Sites: S101 S106 T96	MOTIFS
21	7512289CD1	229	Cytosolic domain: M1-R51 Transmembrane domain: A52-Y74 Non-cytosolic domain: L75-S229	TMHMMER
			PROTEIN INTEGRAL MEMBRANE TRANSMEMBRANE SIGNAL ANCHOR 2A E25 2B E316 E25B PD023945: M1-L156, P108-C228	BLAST_PRODOM
			Potential Phosphorylation Sites: S22 T114 T169 T191 T199 Y192	MOTIFS
22	7512447CD1	1163	signal_cleavage: M1-G17	SPSCAN
			Signal Peptide: M1-G17	HMMER
			Signal Peptide: M1-G20	HMMER
			Signal Peptide: M1-G22	HMMER

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
			Fibronectin type 3 domain: L608-F687, P492-G577	HMMER_SMART
			Immunoglobulin: P143-L224, P233-L320, V26-T131, P408-L488	HMMER_SMART
			Immunoglobulin C-2 Type: Q149-G213, E415-A477, N239-L308, R32-Q115	HMMER_SMART
			Fibronectin type III domain: P492-S580, L608-S692	HMMER_PFAM
			Immunoglobulin domain: L151-A208, G417-A472, S241-P303, I326-P381, G34-V110	HMMER_PFAM
			I type Ig domains from SCOP: P135-L224, G225-I326, A400-V495	HMMER_INCY
			Ig superfamily from SCOP: Q139-P227, I229-P310, I405-V494, S312-F404, G22-F138	HMMER_INCY
			Cytosolic domain: N745-L1163 Transmembrane domain: G722-L744 Non-cytosolic domain: M1-A721	TMHMMER
			Potential Phosphorylation Sites: S74 S209 S285 S585 S700 S797 S819 S850 S960 S1010 S1044 S1059 S1065 S1092 T155 T169 T190 T291 T619 T661 T768 T942 T987 T1064 T1107 T1120 Y106	MOTIFS
			Potential Glycosylation Sites: N124 N188 N235 N239 N256 N452 N497 N508 N696 N1063	MOTIFS

Table 8

SEQ ID NO:	PID	EST ID	SNP ID	EST SNP	CB1 SNP	EST Allele	Allele 1	Allele 2	Amino Acid	Caucasian Allele 1 frequency	African Allele 1 frequency	Asian Allele 1 frequency	Hispanic Allele 1 frequency
28	7490048	3762473H1	SNP00109796	234	854	T	T	C	L252	0.73	n/a	n/a	n/a
29	2133585	2552009H1	SNP00002569	149	1811	G	G	C	noncoding	n/a	n/a	n/a	n/a
30	7509063	5641659H1	SNP00012015	242	3092	C	C	T	S995	0.90	n/d	n/d	0.97
31	5496710	5450982H1	SNP00008406	167	4138	G	A	G	noncoding	n/a	n/a	n/a	n/a
32	72150826	3349030H1	SNP00028258	132	2700	C	C	A	P839	n/a	n/a	n/a	n/a
33	6799476	1667323H1	SNP00003249	208	817	T	C	T	V213	n/a	n/a	n/a	n/a
34	7509488	1667323H1	SNP00003249	208	789	T	C	T	V204	n/a	n/a	n/a	n/a
35	7510060	7091751H1	SNP00027969	189	1003	G	G	A	L289	n/d	n/d	n/d	n/d
37	7510385	181454H1	SNP00016645	144	910	G	G	C	W283	n/d	n/d	n/d	n/d
38	7511618	2246326H1	SNP00003003	219	1301	G	G	A	S321	0.86	n/d	0.97	0.99
40	7506689	2504535H1	SNP00154755	36	2242	T	T	C	noncoding	n/a	n/a	n/a	n/a
41	7510185	4710260H1	SNP00112063	200	2248	G	T	G	V750	0.45	0.35	0.77	0.45
43	7512289	085202H1	SNP00112417	14	298	C	C	A	P31	n/a	n/a	n/a	n/a
44	7512447	2810195T6	SNP00006851	158	3802	A	A	G	noncoding	n/a	n/a	n/a	n/a

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
23/7506758CB1/ 2208	1-601, 189-595, 189-600, 280-600, 316-586, 316-852, 316-857, 316-1977, 344-600, 489-1165, 599-952, 600-671, 603-1225, 616-1183, 684-1284, 705-1159, 716-1207, 730-1014, 736-793, 757-1049, 772-1374, 861-1105, 861-1122, 864-1119, 877-1113, 892-1166, 893-1163, 1101-1385, 1190-1423, 1215-1474, 1231-1487, 1452-1959, 1531-1765, 1533-1976, 1550-1978, 1565-1975, 1633-1928, 1665-1967, 1706-1939, 1708-1914, 1708-1943, 1708-1985, 1741-1972, 1838-1951, 1927-2208, 1975-2173, 1975-2203, 1975-2208, 1985-2076, 1986-2208, 2000-2203, 2149-2208
24/4381669CB1/ 7611	1-321, 227-715, 237-498, 247-723, 275-701, 278-671, 281-697, 286-708, 287-688, 298-875, 298-878, 298-884, 300-679, 309-702, 311-704, 313-698, 320-602, 323-642, 324-626, 329-608, 343-592, 362-701, 374-602, 402-953, 427-712, 491-592, 526-1041, 619-1364, 783-1400, 832-1389, 964-1181, 980-1353, 1086-1448, 1131-1413, 1173-1436, 1207-1850, 1237-1438, 1401-1661, 1415-1532, 1444-1643, 1450-2113, 1543-1613, 1611-3712, 1807-2435, 1929-2435, 2029-2435, 2138-2435, 2277-2397, 2284-2863, 2314-2863, 3291-3666, 3297-3666, 3333-3490, 3471-3858, 3559-3810, 3571-3650, 3808-7611
25/7503554CB1/ 2259	1-599, 1-628, 1-652, 1-666, 1-697, 1-749, 1-771, 1-817, 1-861, 1-875, 1-2214, 22-282, 22-586, 22-587, 22-615, 65-346, 88-616, 96-777, 102-361, 102-371, 107-651, 107-652, 134-673, 153-402, 175-607, 179-868, 190-841, 291-550, 320-707, 334-635, 334-766, 356-870, 385-874, 388-673, 403-1050, 435-678, 448-1105, 473-783, 473-785, 474-688, 481-822, 488-982, 513-816, 513-827, 515-1168, 546-946, 561-988, 578-851, 584-835, 584-1082, 593-1235, 597-871, 654-1347, 656-928, 658-1296, 725-1423, 742-1409, 743-1423, 766-905, 771-1022, 773-1400, 782-1365, 790-1129, 800-1044, 824-1082, 832-1475, 844-1108, 848-1371, 871-1508, 872-1424, 894-1521, 902-1540, 918-1426, 919-1231, 919-1570, 934-1431, 938-1547, 983-1514, 984-1629, 989-1532, 1007-1601, 1023-1589, 1053-1571, 1057-1707, 1157-1907, 1178-1965, 1185-1494, 1185-1667, 1216-1469, 1290-1556, 1293-1545, 1320-1801, 1321-1773, 1322-1722, 1335-1563, 1358-1953, 1362-1595, 1373-1649, 1403-2205, 1411-2187, 1415-2189, 1415-2203, 1422-1710, 1445-2127, 1455-2083, 1468-1713, 1481-2148, 1495-1785, 1533-1786, 1533-1797,
	1533-2116, 1566-2148, 1567-2160, 1577-2053, 1581-2226, 1582-2196, 1585-2155, 1588-2133, 1595-1837, 1606-2157, 1608-1852, 1631-2042, 1653-2159, 1666-2146, 1667-2207, 1691-2159, 1694-2159, 1704-2147, 1718-2159, 1731-2146, 1758-2146, 1773-2042, 1774-2147, 1782-2146, 1783-2146, 1795-2136, 1796-2049, 1805-2159, 1808-2159, 1811-2035, 1838-2140, 1842-2146, 1864-2150, 1873-2149, 1886-2259, 1931-2146, 1963-2253, 1977-2184, 1996-2259, 2051-2258

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
26/7506077CB1/ 1522	1-248, 1-250, 1-1522, 6-268, 12-264, 55-543, 70-350, 72-337, 73-379, 74-340, 74-479, 74-526, 74-546, 78-295, 78-311, 78-349, 79-315, 79-370, 106-398, 123-392, 123-395, 144-413, 177-309, 177-378, 177-394, 177-403, 177-409, 177-410, 177-415, 177-422, 177-427, 177-433, 177-445, 177-447, 177-454, 177-455, 177-479, 177-540, 178-407, 178-513, 178-660, 178-684, 178-699, 178-746, 178-810, 181-300, 181-411, 189-334, 192-730, 194-392, 200-517, 204-906, 223-490, 225-493, 226-547, 229-812, 239-351, 261-547, 295-545, 300-419, 316-507, 318-547, 377-532, 379-547, 491-817, 514-657, 545-720, 554-1106, 556-1138, 557-1064, 564-1231, 568-848, 578-1087, 583-1304, 588-861, 595-1204, 601-1121, 614-1096, 615-1102, 617-1195, 623-1037, 625-904, 628-800, 632-1255, 646-852, 651-847, 656-917, 662-1032, 663-903, 669-1284, 679-1261, 690-970, 707-1192, 712-1263, 715-818, 723-1108, 734-1033, 736-1318, 742-996, 757-988, 758-1522, 776-1036, 787-1078, 796-1441, 799-1364, 802-1085, 812-1008, 813-1064, 813-1070, 813-1436, 831-1294, 832-1261, 833-1456, 835-1102, 837-1505, 842-1084, 851-1113, 851-1522, 852-1144, 854-1080, 860-1110, 866-1121, 866-1162, 866-1513, 867-1106, 867-1167, 868-1105, 869-1151,
	884-1075, 884-1442, 889-1243, 895-1508, 904-1462, 905-1062, 907-1145, 909-1097, 912-1522, 915-1291, 919-1433, 922-1396, 924-1506, 926-1441, 926-1501, 928-1522, 931-1475, 938-1201, 948-1522, 966-1522, 972-1345, 977-1301, 977-1522, 979-1253, 984-1187, 984-1522, 986-1522, 987-1508, 988-1265, 989-1522, 990-1509, 993-1141, 1012-1522, 1013-1265, 1015-1344, 1016-1301, 1017-1125, 1017-1276, 1017-1522, 1019-1521, 1023-1520, 1027-1456, 1028-1231, 1035-1279, 1037-1295, 1037-1301, 1044-1522, 1045-1314, 1045-1522, 1047-1311, 1047-1506, 1047-1522, 1048-1522, 1049-1292, 1051-1226, 1052-1331, 1054-1522, 1056-1513, 1057-1522, 1058-1356, 1060-1522, 1062-1516, 1063-1522, 1069-1511, 1072-1522, 1074-1286, 1074-1511, 1080-1362, 1081-1522, 1084-1513, 1087-1369, 1092-1406, 1093-1513, 1093-1522, 1102-1274, 1103-1513, 1106-1522, 1109-1371, 1109-1504, 1109-1519, 1110-1306, 1110-1352, 1111-1514, 1112-1513, 1115-1469, 1118-1514, 1126-1522, 1130-1518, 1134-1351, 1144-1514, 1145-1513, 1149-1513, 1154-1513, 1155-1518, 1156-1422, 1157-1522, 1162-1516, 1163-1415, 1164-1509, 1166-1512, 1168-1512, 1168-1513, 1168-1514, 1168-1516, 1172-1522, 1174-1508,
	1175-1442, 1176-1513, 1179-1513, 1180-1513, 1182-1511, 1184-1513, 1186-1522, 1187-1512, 1196-1516, 1208-1513, 1208-1516, 1213-1443, 1218-1467, 1220-1514, 1221-1414, 1222-1517, 1223-1474, 1228-1365, 1230-1513, 1233-1513, 1235-1353, 1235-1512, 1237-1441, 1238-1512, 1238-1513, 1242-1441, 1250-1522, 1254-1513, 1261-1514, 1282-1513, 1287-1410, 1292-1515, 1301-1516, 1317-1520, 1328-1520, 1364-1522, 1397-1522, 1413-1522, 1416-1513, 1428-1514

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
27/7506765CB1/ 1099	1-692, 1-748, 1-768, 1-802, 1-808, 1-817, 1-834, 1-909, 2-1099, 3-900, 4-694, 18-970, 111-970, 237-970, 252-970, 307-970, 331-973, 378-970, 451-970, 533-771, 572-815, 572-1099, 822-1051
28/7490048CB1/ 3311	1-254, 1-3244, 155-501, 155-591, 155-606, 155-641, 155-798, 342-505, 342-507, 403-1105, 542-823, 542-869, 542-996, 542-1089, 542-1142, 542-1145, 556-1145, 560-827, 560-999, 667-827, 695-1103, 739-810, 803-1145, 878-1145, 888-1145, 896-1085, 902-1145, 1015-1145, 1039-1145, 1040-2514, 1044-1145, 1046-1145, 1097-1145, 1359-1623, 2689-2753, 2689-3311, 2972-3311, 3098-3311, 3180-3311
29/2133585CB1/ 1874	1-187, 1-271, 1-508, 13-330, 147-815, 304-740, 306-747, 387-855, 393-807, 409-898, 410-898, 432-1045, 547-1218, 604-1241, 615-1874, 634-1289, 637-1254, 681-1207, 730-1442, 779-1356, 791-1364, 796-1364, 810-1448, 812-1227, 824-1296, 844-1320, 855-1287, 857-1370, 887-1363, 902-1344, 930-1420, 930-1430, 1001-1422, 1013-1631, 1051-1433, 1079-1586, 1085-1573, 1410-1874
30/7509063CB1/ 4211	1-232, 1-278, 1-495, 1-4187, 134-733, 203-754, 217-511, 269-935, 311-1123, 394-937, 425-553, 471-929, 482-1158, 520-984, 591-1034, 685-953, 685-1253, 685-1293, 746-1336, 760-1397, 839-1423, 918-1135, 961-1536, 968-1493, 968-1558, 994-1635, 1007-1686, 1095-1669, 1131-1727, 1236-1463, 1360-2012, 1394-1704, 1396-1928, 1397-1693, 1418-1901, 1444-1609, 1477-1986, 1477-2002, 1478-2134, 1536-2068, 1538-1710, 1538-1934, 1583-1856, 1653-2099, 1774-2299, 1836-2374, 1898-2190, 1911-2643, 1955-2517, 2011-2204, 2011-2489, 2011-2516, 2011-2594, 2011-2601, 2061-2238, 2069-2593, 2078-2351, 2101-2361, 2112-2432, 2147-2887, 2168-2599, 2176-2418, 2199-2473, 2212-2950, 2218-2333, 2234-2811, 2237-2875, 2244-2851, 2245-2773, 2295-2537, 2297-2548, 2346-2993, 2351-2682, 2353-3000, 2356-2795, 2389-2915, 2393-2658, 2397-2785, 2402-2687, 2422-2701, 2427-3029, 2434-3089, 2440-3079, 2468-3025, 2474-2922, 2477-2698, 2486-3059, 2491-2781, 2492-2724, 2492-3033, 2503-2760, 2525-2664, 2531-2760, 2545-3390, 2549-2991, 2558-2802, 2563-3283, 2618-2856, 2634-2903, 2636-2851, 2641-3469, 2645-2979, 2646-2905, 2693-3046, 2713-2978, 2725-3173, 2740-3354,

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
30 cont	2759-3047, 2759-3237, 2775-3262, 2779-3064, 2779-3075, 2798-3029, 2799-3389, 2806-3026, 2806-3074, 2806-3087, 2806-3155, 2820-3118, 2820-3151, 2823-2961, 2826-3104, 2834-3087, 2837-3115, 2849-3196, 2850-3114, 2853-3025, 2853-3448, 2856-3090, 2858-3090, 2868-3434, 2890-3405, 2899-3134, 2902-3099, 2902-3150, 2911-3205, 2912-3166, 2915-3137, 2916-3529, 2917-3396, 2938-3212, 2944-3384, 2944-3599, 2946-3218, 2956-3217, 2961-3446, 2967-3200, 2983-3253, 2999-3629, 2999-3636, 3012-3218, 3023-3278, 3030-3306, 3030-3632, 3041-3234, 3043-3218, 3094-3380, 3095-3673, 3115-3421, 3122-3762, 3130-3430, 3132-3716, 3145-3404, 3145-3643, 3148-3675, 3155-3720, 3156-3605, 3156-3688, 3156-3734, 3197-3747, 3199-3735, 3202-3810, 3203-3887, 3234-3605, 3239-3605, 3267-3548, 3288-3572, 3289-3574, 3293-3828, 3307-3587, 3323-3796, 3328-3531, 3330-3636, 3344-3607, 3346-3617, 3352-3619, 3355-3808, 3360-3638, 3361-3825, 3363-3553, 3363-3635, 3363-3636, 3374-4035, 3389-4044, 3406-3835, 3408-3725, 3414-4023, 3416-3666, 3422-3682, 3422-3915, 3437-3701, 3453-3740, 3455-3699, 3456-3717, 3456-3720, 3471-3698, 3473-4098, 3478-3736, 3481-3954, 3486-3938,
	3494-3783, 3495-3752, 3498-3722, 3499-3769, 3512-4105, 3514-4048, 3515-3768, 3522-3795, 3530-4144, 3543-4119, 3549-3882, 3561-4141, 3582-4195, 3589-3812, 3589-4165, 3591-4119, 3591-4211, 3594-4090, 3596-3878, 3601-3877, 3603-3817, 3603-3917, 3608-3885, 3615-4105, 3619-3876, 3627-3774, 3630-4094, 3631-3882, 3631-3979, 3632-3792, 3632-3901, 3634-4105, 3639-3873, 3640-4148, 3640-4208, 3641-4094, 3642-3949, 3644-4081, 3646-3859, 3646-3872, 3648-4097, 3651-3875, 3652-3897, 3652-4103, 3663-4060, 3665-4094, 3671-3939, 3671-3982, 3674-4181, 3677-3956, 3677-4078, 3683-4095, 3686-3930, 3687-4094, 3689-4205, 3691-4211, 3695-4152, 3706-4146, 3708-3958, 3716-3932, 3717-3992, 3717-3993, 3719-4007, 3721-4182, 3723-4186, 3724-3867, 3729-4006, 3732-4186, 3734-4186, 3737-4186, 3738-4035, 3738-4205, 3739-4103, 3739-4204, 3740-4191, 3742-4186, 3743-4056, 3743-4181, 3749-4150, 3751-4204, 3752-4186, 3754-4012, 3754-4186, 3758-4208, 3764-3947, 3764-4182, 3765-4180, 3768-4186, 3770-3955, 3772-4012, 3785-4178, 3788-4049, 3792-4179, 3793-4091, 3799-4204, 3805-4187, 3806-4094, 3825-4181, 3833-4095, 3839-4105, 3841-4103, 3844-4058, 3845-4127,
	3855-4146, 3859-4185, 3860-4158, 3870-4186, 3880-4179, 3881-4179, 3889-4114, 3904-4148, 3906-4186, 3906-4201, 3953-4186, 3954-4181, 3960-4186, 3960-4187, 3967-4093, 3986-4194, 3995-4186, 4010-4208, 4028-4147, 4028-4186, 4029-4205, 4031-4129, 4033-4181, 4073-4186

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
31/5496710CB1/ 5006	1-240, 1-548, 1-2915, 20-548, 81-709, 81-839, 92-548, 103-322, 162-694, 242-2395, 485-1087, 509-1268, 723-1268, 739-1268, 1050-1087, 1467-1600, 1467-1602, 1515-1921, 1515-2191, 1650-1799, 1703-1993, 2012-2041, 2040-2443, 2066-2323, 2073-2395, 2150-2554, 2151-2462, 2263-2502, 2303-2730, 2368-2633, 2443-2716, 2464-3139, 2490-2861, 2735-2913, 2765-2915, 2920-3172, 2920-3260, 3096-3406, 3153-3589, 3180-3463, 3222-3371, 3297-3520, 3301-3967, 3323-3816, 3426-4028, 3429-3663, 3462-3718, 3462-3887, 3481-3985, 3483-3744, 3694-4131, 3746-4046, 3786-4623, 3802-4134, 3810-4400, 3859-4097, 3890-4112, 3906-4168, 3948-4232, 3973-4532, 3975-4384, 3988-4411, 4001-4601, 4173-4633, 4187-4633, 4190-4633, 4328-4638, 4354-4380, 4354-4384, 4354-4385, 4355-4385, 4359-4562, 4359-4564, 4359-4645, 4375-4589, 4380-4735, 4410-5006, 4414-4633, 4416-4633
32/72150826CB1/ 4125	1-592, 2-4125, 6-546, 274-601, 421-1013, 428-1017, 607-1318, 654-1026, 658-731, 658-1030, 686-1301, 718-995, 722-1142, 726-1148, 730-984, 733-900, 733-984, 808-1368, 808-1372, 835-1362, 838-1366, 898-1276, 1035-1557, 1039-1549, 1069-1877, 1069-3791, 1108-1868, 1160-1860, 1543-2038, 1543-2048, 1543-2075, 1543-2078, 1546-2174, 1701-2522, 1736-2371, 1824-2510, 1851-2519, 1851-2545, 1851-2657, 1851-2701, 1851-2742, 1880-2709, 1890-2733, 1899-2545, 1983-2823, 2013-2834, 2019-2578, 2064-2693, 2069-2650, 2081-2776, 2089-2608, 2122-2711, 2148-2795, 2178-2694, 2217-2775, 2219-2811, 2247-2808, 2287-3021, 2292-3058, 2321-2886, 2369-2833, 2371-3158, 2376-3190, 2388-3179, 2411-2909, 2415-2971, 2422-2940, 2462-3200, 2499-3066, 2505-3121, 2516-3041, 2519-3140, 2525-3280, 2536-3243, 2539-3136, 2542-3186, 2554-3260, 2572-3064, 2585-3387, 2595-3219, 2648-3276, 2653-3452, 2656-3175, 2681-3418, 2687-3331, 2688-3415, 2701-3447, 2705-3555, 2713-3380, 2714-3180, 2720-2951, 2720-3164, 2729-3364, 2737-3472, 2740-3175, 2747-3204, 2766-3559, 2767-3420, 2774-3321, 2780-3129, 2780-3282, 2780-3332, 2780-3340, 2780-3385, 2780-3397, 2780-3452, 2800-3486,
	2829-3458, 2857-3444, 2860-3726, 2876-3491, 2884-3685, 2893-3611, 2894-3473, 2896-3190, 2899-3736, 2906-3516, 2912-3427, 2920-3537, 2923-3614, 2925-3429, 2948-3587, 2953-3672, 2970-3241, 2970-3433, 2983-3561, 2997-3543, 3010-3749, 3022-3673, 3052-3618, 3057-3700, 3057-3707, 3061-3711, 3082-3626, 3083-3780, 3096-3621, 3099-3649, 3114-3773, 3128-3685, 3586-3832, 3586-4072, 3586-4125, 3653-3839, 3653-3864, 3653-3970, 3694-3948, 3731-4125, 3740-3989, 3787-4085, 3844-4125

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
33/6799476CB1/ 3597	1-509, 1-799, 231-856, 259-861, 313-619, 357-851, 378-930, 405-527, 405-694, 405-697, 405-705, 416-711, 432-679, 440-1080, 441-975, 459-819, 472-1164, 503-944, 503-947, 503-1085, 533-806, 538-834, 538-1165, 543-1167, 546-1094, 556-830, 559-1014, 566-1050, 576-775, 579-1167, 583-1058, 584-1165, 588-1165, 596-880, 606-1130, 606-1165, 610-841, 619-808, 619-1228, 626-1013, 626-1111, 626-1114, 629-900, 637-1136, 640-894, 640-1132, 642-768, 649-959, 667-1136, 668-1136, 670-948, 673-903, 673-1113, 673-1136, 676-1136, 676-1151, 679-780, 679-1136, 681-1136, 683-1136, 683-1165, 683-1170, 687-1136, 687-1151, 689-1114, 690-1151, 692-1136, 694-1136, 696-1136, 699-1136, 701-1136, 701-1164, 708-1155, 712-961, 714-1118, 720-1136, 722-1155, 745-1163, 747-1006, 747-1007, 755-1136, 758-1136, 759-1136, 764-1011, 774-1136, 817-1375, 827-1136, 827-1552, 830-1362, 830-1426, 832-1353, 836-1374, 840-1418, 846-1497, 855-1426, 869-1604, 870-1420, 872-1381, 881-1478, 886-1426, 890-1415, 913-1605, 953-1634, 955-1363, 960-1535, 974-1240, 988-1252, 1013-1621, 1032-1545, 1043-1597, 1053-1222, 1075-1505, 1089-1613, 1156-2039, 1175-1844, 1252-1456, 1255-1684, 1331-1954,
	1374-1899, 1374-1901, 1437-2275, 1479-1883, 1738-2015, 1738-2186, 1749-2007, 1760-2050, 1768-1968, 1783-1982, 1818-2059, 1818-2387, 1929-2190, 1947-2249, 1995-2255, 2002-2244, 2002-2407, 2018-2232, 2041-2291, 2096-2298, 2096-2321, 2096-2506, 2125-2337, 2125-2384, 2142-2359, 2145-2447, 2149-2325, 2152-2388, 2169-2429, 2195-2452, 2212-2452, 2291-2506, 2291-2763, 2352-2615, 2352-2886, 2368-2432, 2374-2982, 2377-3055, 2378-3198, 2379-2625, 2391-3198, 2423-2977, 2430-3077, 2434-3016, 2437-2667, 2458-3009, 2512-2982, 2519-3098, 2531-3005, 2542-2763, 2542-2979, 2543-3032, 2547-3271, 2555-3222, 2557-3045, 2562-3175, 2564-2790, 2567-3142, 2578-3011, 2584-3043, 2584-3187, 2594-3218, 2605-3198, 2613-3208, 2626-3208, 2627-3330, 2635-2924, 2638-2878, 2644-2980, 2649-3323, 2656-3108, 2669-3208, 2677-2927, 2682-3004, 2682-3225, 2684-3205, 2687-3328, 2692-2893, 2692-3114, 2698-3320, 2712-3120, 2714-2978, 2715-3315, 2719-3193, 2724-2990, 2737-2951, 2760-3398, 2767-3253, 2767-3371, 2770-3010, 2780-3075, 2781-3030, 2782-3144, 2788-3293, 2788-3456, 2790-3245, 2791-3045, 2791-3286, 2793-3533, 2794-3447, 2804-3331, 2805-3046, 2808-2969,
	2812-3025, 2813-3383, 2823-3460, 2830-3048, 2831-3400, 2832-3400, 2834-3137, 2849-3377, 2849-3446, 2851-3362, 2851-3440, 2854-3473, 2867-3396, 2876-3446, 2879-3178, 2884-3505, 2888-3142, 2889-3080, 2889-3447, 2891-3448, 2892-3438, 2894-3012, 2894-3172, 2894-3444, 2894-3449, 2896-3597, 2900-3122, 2902-3192, 2906-3155, 2906-3590, 2909-3163, 2909-3460, 2911-3147, 2911-3153, 2911-3399, 2919-3014, 2919-3117, 2922-3181, 2923-3417, 2924-3165, 2927-3167, 2927-3211, 2932-3160, 2935-3188, 2937-3210, 2944-3183, 2954-3203, 2961-3266

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
34/7509488CB1/ 3569	1-567, 47-622, 221-487, 255-514, 271-723, 271-787, 298-781, 328-796, 355-806, 373-807, 393-481, 393-499, 393-591, 393-666, 393-669, 393-677, 393-683, 393-771, 393-823, 393-828, 393-833, 396-902, 396-3569, 404-651, 412-1052, 413-947, 431-791, 444-1136, 475-613, 475-916, 475-919, 475-1057, 505-778, 510-806, 510-1137, 515-1139, 518-1066, 528-802, 531-986, 538-1022, 548-747, 548-861, 551-1139, 555-1030, 556-1137, 560-1137, 568-852, 578-1102, 578-1137, 582-813, 591-780, 591-1200, 598-985, 598-1083, 598-1086, 601-872, 609-1108, 612-866, 612-1104, 614-740, 615-869, 621-931, 639-1108, 640-1108, 642-920, 645-875, 645-1085, 645-1108, 648-1108, 648-1123, 651-752, 651-1108, 653-1108, 655-1108, 659-1108, 659-1123, 662-1123, 664-1108, 666-1108, 668-1108, 671-1108, 673-1108, 680-1127, 684-933, 686-1090, 692-1108, 694-1127, 719-978, 719-979, 727-1108, 730-1108, 731-1108, 746-1108, 789-1347, 799-1108, 799-1524, 802-1334, 802-1398, 804-1325, 808-1346, 812-1390, 818-1469, 827-1398, 841-1576, 842-1392, 844-1353, 853-1450, 858-1398, 862-1387, 885-1577, 925-1606, 927-1335, 932-1507, 946-1212, 960-1224, 985-1593, 1004-1517, 1015-1569, 1025-1194, 1047-1477, 1061-1585,
	1128-2011, 1147-1816, 1224-1428, 1227-1656, 1303-1926, 1346-1871, 1346-1873, 1409-2247, 1451-1855, 1710-1987, 1710-2158, 1721-1979, 1732-2022, 1740-1940, 1755-1954, 1790-2031, 1790-2359, 1901-2162, 1919-2221, 1967-2227, 1974-2216, 1974-2379, 1990-2204, 2013-2263, 2068-2270, 2068-2293, 2068-2478, 2097-2309, 2097-2356, 2114-2331, 2117-2419, 2121-2297, 2124-2360, 2141-2401, 2167-2424, 2184-2424, 2263-2478, 2263-2735, 2324-2587, 2324-2858, 2340-2404, 2346-2954, 2349-3027, 2350-3170, 2351-2597, 2363-3170, 2395-2949, 2402-3049, 2406-2988, 2409-2639, 2430-2981, 2484-2954, 2491-3070, 2503-2977, 2514-2735, 2514-2951, 2515-3004, 2519-3243, 2527-3194, 2529-3017, 2534-3147, 2536-2762, 2539-3114, 2550-2983, 2556-3015, 2556-3159, 2566-3190, 2577-3170, 2585-3180, 2598-3180, 2599-3302, 2607-2896, 2610-2850, 2616-2952, 2621-3295, 2628-3080, 2641-3180, 2649-2899, 2654-2976, 2654-3197, 2656-3177, 2659-3300, 2664-2865, 2664-3086, 2670-3292, 2684-3092, 2686-2950, 2687-3287, 2691-3165, 2696-2962, 2709-2923, 2732-3370, 2739-3225, 2739-3343, 2742-2982, 2752-3047, 2753-3002, 2754-3116, 2760-3265, 2760-3428, 2762-3217, 2763-3017, 2763-3258,
	2765-3505, 2766-3419, 2776-3303, 2777-3018, 2780-2941, 2784-2997, 2785-3355, 2795-3432, 2802-3020, 2803-3372, 2804-3372, 2806-3109, 2821-3349, 2821-3418, 2823-3334, 2823-3412, 2826-3445, 2839-3368, 2848-3108, 2848-3418, 2851-3150, 2856-3477, 2859-3106, 2860-3114, 2861-3052, 2861-3419, 2863-3420, 2864-3410, 2866-2984, 2866-3144, 2866-3416, 2866-3421, 2868-3569, 2872-3094, 2874-3164, 2878-3127, 2878-3562, 2881-3135, 2881-3432, 2883-3119, 2883-3125, 2883-3371, 2891-2986, 2891-3089, 2894-3153, 2895-3389, 2896-3137, 2899-3139, 2899-3183, 2904-3132, 2907-3160, 2909-3182, 2916-3155, 2926-3175, 2933-3238

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
35/7510060CB1/ 2431	1-530, 12-268, 12-640, 14-251, 14-566, 14-579, 14-632, 14-653, 14-686, 14-697, 17-572, 25-343, 26-579, 26-657, 26-672, 26-693, 26-716, 27-296, 27-437, 27-685, 27-876, 27-2426, 30-320, 31-234, 33-339, 34-551, 37-599, 39-635, 41-275, 42-352, 43-587, 43-714, 43-726, 43-801, 43-840, 43-853, 43-854, 43-857, 43-867, 44-561, 47-721, 48-302, 50-479, 239-890, 381-983, 414-997, 421-680, 421-692, 465-569, 520-1196, 566-839, 569-832, 569-1146, 574-1221, 591-1202, 712-1329, 733-1049, 733-1118, 786-1062, 800-1011, 832-1373, 856-1117, 885-1400, 889-1465, 928-1148, 973-1473, 1078-1371, 1120-1477, 1152-1352, 1153-1644, 1159-1426, 1165-1760, 1168-1789, 1192-1476, 1193-1473, 1197-1657, 1210-1512, 1233-1480, 1248-1568, 1261-1506, 1331-1582, 1397-1516, 1399-1953, 1406-1696, 1409-1700, 1416-1645, 1416-1664, 1428-2344, 1438-1681, 1455-1711, 1477-1553, 1480-1707, 1523-1781, 1544-2344, 1572-2344, 1737-1990, 1737-1991, 1760-2424, 1762-1827, 1796-2286, 1843-2357, 1950-2351, 1950-2393, 1954-2403, 1965-2376, 2000-2431, 2013-2426, 2059-2309, 2087-2415, 2098-2415, 2099-2417, 2100-2412, 2110-2412, 2112-2413, 2177-2427, 2216-2415, 2227-2415, 2238-2415, 2287-2426, 2299-2406, 2299-2426
36/7510226CB1/ 3842	1-442, 1-3737, 802-1028, 869-1196, 1090-1337, 1117-1318, 1473-2148, 1478-2258, 1478-2284, 1481-2279, 1549-2159, 1556-1777, 1631-2169, 1632-2077, 1632-2166, 1641-2091, 1643-2347, 1648-2086, 1649-2268, 1653-2236, 1663-2374, 1664-2442, 1679-2167, 1706-2185, 1727-2327, 1734-2437, 1735-2148, 1738-2281, 1756-2330, 1756-2333, 1757-2284, 1762-2409, 1762-2488, 1781-2405, 1791-2563, 1792-2215, 1793-2054, 1793-2083, 1802-2112, 1804-2276, 1808-2200, 1809-2312, 1825-2504, 1825-2704, 1827-2362, 1833-2522, 1839-2125, 1856-2484, 1863-2541, 1872-2394, 1885-2064, 1885-2224, 1936-2513, 1939-2220, 1951-2213, 1952-2608, 1957-2220, 1963-2700, 1977-2548, 1980-2238, 1980-2450, 1989-2483, 2002-2703, 2026-2426, 2026-2738, 2029-2615, 2050-2813, 2066-2783, 2070-2767, 2072-2312, 2107-2727, 2108-2419, 2109-2616, 2119-2338, 2122-2360, 2134-2394, 2142-2689, 2146-2829, 2147-2801, 2159-2322, 2167-2286, 2176-2415, 2179-2542, 2182-2445, 2193-2887, 2214-2815, 2216-2782, 2218-2468, 2219-2455, 2231-2866, 2245-2555, 2270-2567, 2275-2968, 2293-2520, 2294-2499,

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
36 cont	2294-2892, 2297-2520, 2306-2459, 2311-2583, 2312-2857, 2317-2882, 2331-2899, 2354-2772, 2355-2583, 2359-2581, 2375-2588, 2378-3008, 2386-2586, 2395-2637, 2410-2673, 2410-2681, 2412-2960, 2416-2939, 2417-2945, 2417-3010, 2420-3096, 2423-2716, 2423-2962, 2444-2711, 2445-3193, 2464-2888, 2471-2706, 2481-2759, 2490-2734, 2494-2753, 2494-2758, 2494-3154, 2497-3112, 2500-2741, 2501-2799, 2508-2963, 2509-2790, 2510-2760, 2510-2818, 2511-2701, 2512-3112, 2531-3112, 2533-3083, 2537-2959, 2548-3033, 2552-2827, 2552-3194, 2571-2835, 2572-3197, 2577-3201, 2591-2837, 2593-2814, 2598-2875, 2607-3112, 2608-3112, 2616-3041, 2619-2872, 2619-2892, 2620-2917, 2622-2906, 2627-3524, 2636-2903, 2636-2914, 2642-2793, 2645-2746, 2645-2929, 2655-3112, 2658-3179, 2658-3290, 2661-3175, 2682-3229, 2684-3171, 2686-2969, 2693-2903, 2695-3170, 2698-2964, 2698-3112, 2700-2958, 2700-3170, 2700-3178, 2711-2973, 2711-3175, 2717-2953, 2718-3420, 2722-3314, 2730-2986, 2733-2947, 2735-3183, 2737-3241, 2738-3104, 2744-3195, 2762-3223, 2763-3414, 2765-3223,
	2770-3059, 2770-3319, 2771-3054, 2771-3480, 2775-3460, 2779-3050, 2789-3049, 2837-2901, 2867-3137, 2868-3196, 2870-3655, 2877-3150, 2880-3136, 2880-3279, 2884-3173, 2886-3158, 2886-3167, 2895-2983, 2895-3143, 2897-3191, 2900-3507, 2910-3463, 2917-3398, 2934-3294, 2935-3198, 2943-3647, 2947-3008, 2948-3177, 2950-3153, 2951-3212, 2962-3217, 2962-3279, 2963-3472, 2963-3695, 2969-3246, 2969-3630, 2971-3250, 2971-3287, 2971-3514, 2972-3410, 2972-3636, 2976-3554, 2980-3571, 2990-3248, 2997-3275, 2997-3283, 3000-3655, 3003-3467, 3004-3228, 3008-3625, 3011-3468, 3012-3290, 3013-3273, 3015-3446, 3018-3274, 3020-3410, 3028-3572, 3029-3523, 3030-3424, 3033-3300, 3034-3251, 3042-3295, 3048-3608, 3051-3488, 3051-3653, 3052-3652, 3053-3488, 3057-3644, 3058-3660, 3062-3632, 3064-3269, 3071-3309, 3071-3433, 3071-3486, 3072-3449, 3073-3663, 3075-3374, 3077-3196, 3077-3548, 3085-3472, 3086-3568, 3098-3719, 3100-3415, 3100-3533, 3103-3709, 3105-3659, 3108-3361, 3108-3392, 3108-3572, 3112-3704, 3118-3392, 3121-3519, 3124-3721, 3129-3285,

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
36 cont	3132-3598, 3138-3319, 3153-3726, 3157-3415, 3157-3447, 3157-3449, 3161-3300, 3161-3375, 3161-3410, 3163-3404, 3163-3441, 3163-3700, 3166-3634, 3166-3653, 3169-3728, 3171-3346, 3171-3420, 3171-3443, 3171-3454, 3172-3687, 3175-3455, 3183-3383, 3184-3657, 3184-3715, 3185-3449, 3185-3725, 3185-3727, 3190-3664, 3190-3730, 3199-3454, 3207-3440, 3207-3457, 3208-3702, 3208-3703, 3211-3708, 3212-3457, 3214-3458, 3214-3724, 3226-3664, 3234-3517, 3236-3699, 3238-3478, 3241-3703, 3247-3703, 3248-3414, 3248-3485, 3248-3706, 3250-3666, 3256-3710, 3257-3673, 3258-3642, 3258-3674, 3265-3585, 3265-3728, 3270-3721, 3270-3724, 3272-3708, 3272-3731, 3273-3702, 3273-3704, 3273-3725, 3273-3737, 3274-3701, 3275-3679, 3275-3710, 3276-3703, 3277-3706, 3279-3704, 3279-3758, 3286-3563, 3287-3725, 3293-3567, 3295-3721, 3295-3731, 3297-3511, 3305-3711, 3306-3542, 3306-3638, 3307-3705, 3309-3702, 3310-3702, 3315-3699, 3315-3703, 3317-3710, 3319-3605, 3320-3704, 3321-3723, 3323-3709, 3324-3579, 3324-3703, 3325-3668, 3330-3706, 3332-3709, 3333-3703,
	3334-3705, 3337-3703, 3342-3701, 3345-3626, 3346-3664, 3348-3707, 3351-3549, 3352-3439, 3362-3703, 3384-3703, 3392-3555, 3394-3555, 3406-3709, 3413-3705, 3414-3716, 3415-3512, 3415-3777, 3415-3842, 3420-3769, 3420-3774, 3427-3678, 3427-3687, 3431-3658, 3441-3695, 3447-3703, 3451-3705, 3452-3725, 3455-3703, 3458-3730, 3462-3700, 3464-3703, 3466-3709, 3486-3704, 3491-3749, 3492-3717, 3494-3718, 3514-3702, 3524-3717, 3527-3682, 3534-3753, 3536-3703, 3541-3703, 3547-3698, 3570-3711, 3605-3737, 3605-3739, 3612-3731, 3614-3703, 3638-3708, 3650-3703, 3652-3703
37/7510385CB1/ 2892	1-113, 1-114, 1-247, 1-260, 6-287, 6-2883, 9-292, 12-309, 14-288, 22-332, 26-624, 38-623, 66-771, 67-513, 71-686, 77-487, 84-771, 97-602, 99-644, 125-624, 126-693, 146-577, 148-734, 168-769, 180-759, 181-759, 194-446, 194-455, 211-951, 213-951, 218-478, 235-784, 275-771, 283-771, 287-698, 319-883, 319-960, 345-771, 354-771, 366-469, 403-602, 515-767, 518-1053, 525-767, 552-892, 638-1135, 670-975, 675-944, 677-933, 726-1311, 741-1311, 850-1513, 1027-1273, 1029-1453, 1098-1631, 1127-1363, 1235-1649, 1240-1681, 1258-1405, 1277-1920, 1278-1704, 1288-1848, 1295-1383, 1319-1593, 1330-1634, 1335-1885, 1362-1667, 1380-1625, 1439-1648, 1454-1716, 1454-1731, 1474-1604, 1479-1845, 1496-2043, 1497-1688, 1513-1736, 1518-1709, 1519-1757, 1529-1769, 1529-1805, 1536-1665, 1546-1919, 1551-1859, 1559-1859, 1568-1824, 1574-1847, 1637-2222, 1640-1943, 1651-1975, 1651-1980, 1668-1947, 1674-2288, 1682-1849, 1693-2020, 1693-2495, 1756-1969, 1769-2253, 1779-2283, 1799-2022, 1799-2547, 1831-2354, 1835-2555, 1836-1969, 1840-2223, 1843-2078, 1843-2079, 1854-2115, 1884-2149, 1917-2582, 1919-2197, 1921-2028, 1921-2191, 1921-2289, 1931-2670, 1934-2209, 1943-2371,

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
37 cont	1944-2675, 1945-2433, 1950-2387, 1950-2451, 1983-2619, 1995-2175, 2004-2786, 2004-2816, 2026-2475, 2034-2286, 2043-2530, 2076-2818, 2086-2693, 2089-2878, 2097-2437, 2129-2322, 2129-2404, 2138-2693, 2138-2807, 2140-2803, 2141-2694, 2145-2543, 2159-2481, 2160-2808, 2162-2317, 2168-2708, 2175-2433, 2178-2399, 2182-2864, 2185-2390, 2220-2361, 2222-2361, 2229-2860, 2245-2494, 2246-2556, 2265-2564, 2265-2616, 2269-2883, 2282-2552, 2301-2548, 2301-2589, 2303-2558, 2303-2571, 2325-2883, 2328-2843, 2330-2883, 2335-2883, 2338-2767, 2342-2590, 2348-2606, 2352-2580, 2352-2797, 2353-2545, 2354-2552, 2355-2570, 2359-2611, 2367-2841, 2371-2800, 2396-2856, 2397-2837, 2408-2513, 2409-2892, 2410-2651, 2427-2883, 2432-2880, 2435-2883, 2438-2883, 2446-2860, 2453-2887, 2460-2613, 2460-2643, 2460-2860, 2466-2754, 2467-2879, 2468-2892, 2469-2887, 2470-2860, 2475-2696, 2475-2702, 2477-2843, 2477-2883, 2479-2860, 2483-2879, 2484-2879, 2486-2892, 2493-2857, 2500-2860, 2504-2757, 2507-2859, 2508-2857, 2509-2858, 2512-2880, 2515-2860, 2527-2780, 2529-2804, 2529-2809, 2532-2785, 2545-2823, 2548-2879, 2551-2790, 2555-2787, 2570-2888, 2577-2883,
	2580-2857, 2581-2883, 2584-2886, 2586-2892, 2587-2871, 2590-2857, 2600-2879, 2606-2843, 2606-2860, 2616-2829, 2629-2863, 2633-2874, 2647-2837, 2648-2821, 2649-2857, 2657-2881, 2661-2883, 2664-2853, 2664-2855, 2664-2856, 2664-2870, 2664-2872, 2664-2885, 2666-2879, 2668-2870, 2669-2879, 2670-2856, 2670-2870, 2670-2877, 2670-2879, 2675-2883, 2676-2892, 2677-2860, 2678-2883, 2693-2857, 2698-2883, 2703-2892, 2740-2786, 2743-2879, 2759-2883, 2771-2872, 2771-2883
38/7511618CB1/ 2756	1-424, 1-2756, 46-423, 57-295, 57-310, 57-412, 57-413, 62-415, 98-235, 104-406, 107-424, 132-402, 136-424, 417-518, 417-839, 519-754, 519-765, 519-773, 519-782, 519-783, 519-844, 519-996, 519-1103, 521-972, 541-983, 541-1138, 571-1058, 586-1252, 589-970, 598-948, 598-953, 614-1058, 626-1102, 635-1138, 645-913, 654-1124, 666-1249, 667-1021, 668-926, 668-1031, 674-933, 675-924, 691-1319, 696-953, 718-1188, 727-997, 732-1058, 733-1132, 743-1124, 748-1049, 753-1324, 792-1025, 796-1193, 801-1087, 805-1123, 807-1193, 819-1224, 846-1137, 881-1058, 901-1184, 903-1777, 920-1192, 931-1213, 933-1644, 972-1265, 972-1566, 998-1593, 1008-1275, 1021-1187, 1038-1319, 1046-1491, 1046-1546, 1061-1296, 1083-1335, 1108-1626, 1136-1160, 1136-1281, 1136-1302, 1136-1529, 1137-1581, 1157-1536, 1158-1708, 1182-1434, 1183-1871, 1191-1454, 1195-1717, 1197-1692, 1218-1495, 1218-1609, 1251-1522, 1267-1907, 1272-1563, 1281-1505, 1286-1561, 1323-1654, 1331-1556, 1331-1767, 1350-1592, 1356-1592, 1374-1694, 1377-1898, 1400-1655, 1417-1725, 1425-1653, 1425-1668, 1427-1594, 1433-1818, 1440-1954, 1442-1929, 1459-1741, 1477-1623, 1478-1745, 1483-1939, 1490-1651, 1493-1955,

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
38 cont	1505-1927, 1509-2147, 1510-1954, 1513-1931, 1516-1954, 1519-1806, 1527-1937, 1532-2047, 1534-2068, 1537-2121, 1540-1822, 1543-1953, 1549-1936, 1560-1954, 1564-1954, 1573-1846, 1586-1954, 1596-1937, 1596-1947, 1610-1950, 1616-1794, 1625-1920, 1628-1940, 1629-1922, 1645-1881, 1645-2131, 1646-1886, 1646-1930, 1646-2155, 1667-1928, 1670-1950, 1678-1957, 1693-1953, 1712-1990, 1713-1989, 1723-1935, 1729-1908, 1738-2122, 1746-2414, 1788-2081, 1820-2178, 1820-2297, 1823-1935, 1828-2101, 1828-2298, 1846-2103, 1851-2365, 1870-2148, 1889-2500, 1908-2445, 1920-2492, 1943-2239, 1945-2100, 1957-2193, 1957-2194, 1969-2392, 1969-2436, 1971-2243, 1973-2232, 1987-2274, 1994-2486, 1995-2299, 2014-2299, 2034-2477, 2044-2492, 2047-2477, 2054-2342, 2058-2471, 2060-2278, 2060-2631, 2062-2471, 2068-2485, 2068-2489, 2071-2346, 2072-2335, 2076-2472, 2087-2486, 2090-2666, 2106-2367, 2110-2317, 2110-2450, 2114-2626, 2115-2201, 2122-2684, 2123-2490, 2126-2487, 2149-2669, 2149-2676, 2156-2543, 2162-2457, 2168-2541, 2187-2466, 2187-2626, 2208-2541, 2215-2631, 2226-2487, 2228-2459, 2236-2504, 2307-2680, 2334-2587, 2340-2746,
	2343-2547, 2348-2492, 2348-2614, 2399-2624, 2443-2681, 2475-2756, 2495-2678, 2558-2680
39/6244135CB1/ 873	1-302, 12-348, 14-302, 106-836, 124-413, 165-714, 165-727, 165-771, 206-477, 326-506, 392-639, 447-705, 459-873, 497-827, 499-683, 607-873
40/7506689CB1/ 2660	1-218, 1-411, 1-811, 36-718, 115-408, 148-405, 151-399, 265-408, 407-538, 407-577, 407-611, 407-617, 407-650, 407-675, 407-740, 407-750, 407-787, 407-804, 407-957, 407-2595, 408-1113, 489-1018, 515-1111, 515-1112, 538-813, 558-1111, 572-1067, 577-1089, 581-1108, 602-1111, 603-675, 605-1111, 606-1111, 609-1111, 613-923, 613-1111, 618-1111, 620-956, 620-1083, 620-1166, 620-1236, 624-1111, 671-1111, 674-1111, 715-934, 715-955, 715-1159, 715-1218, 715-1228, 715-1253, 715-1284, 719-1313, 757-1111, 881-1397, 888-1318, 899-1455, 941-1075, 973-1480, 1055-1418, 1086-1632, 1094-1381, 1100-1686, 1148-1702, 1186-1656, 1190-1437, 1196-1429, 1196-1758, 1204-1313, 1222-1749, 1225-1797, 1251-1590, 1349-1748, 1349-1979, 1418-1943, 1420-1954, 1422-1943, 1436-2352, 1456-1925, 1473-2011, 1485-2111, 1491-2352, 1499-2352, 1520-2032, 1532-2352, 1536-1810, 1592-2352, 1612-2191, 1615-2348, 1632-2216, 1635-2125, 1659-2352, 1669-2190, 1677-2352, 1682-2380, 1684-1892, 1700-2352, 1711-2162, 1729-2394, 1738-2353, 1742-2449, 1772-2017, 1776-2398, 1777-2386, 1780-2454, 1784-2203, 1842-2439, 1865-2432, 1909-2454, 1918-2454, 1984-2453, 1986-2454, 1999-2454, 2020-2448,
	2022-2311, 2048-2449, 2051-2660, 2053-2448, 2098-2472, 2135-2450, 2155-2453, 2168-2454, 2207-2438, 2241-2454, 2241-2498, 2386-2454

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
41/7510185CB1/ 2761	1-2759, 103-686, 104-686, 512-784, 512-1040, 685-1492, 796-1296, 1383-1613, 1614-1857, 1614-2035, 1614-2044, 1614-2191, 1614-2240, 1614-2338, 1614-2368, 1614-2374, 1614-2386, 1614-2486, 1678-2262, 1791-2070, 1873-2681, 1888-2165, 1888-2168, 1888-2269, 1888-2344, 1888-2441, 1987-2761, 2026-2524, 2050-2314, 2133-2761, 2195-2761, 2212-2761, 2216-2740, 2242-2457, 2251-2761, 2314-2749, 2315-2745, 2315-2749, 2317-2624, 2317-2656, 2317-2705, 2346-2755, 2371-2761, 2414-2749, 2490-2755
42/1420867CB1/ 1440	1-252, 1-268, 1-311, 1-418, 1-460, 1-511, 1-529, 1-589, 1-605, 1-691, 1-711, 1-732, 1-734, 1-881, 12-814, 44-997, 196-432, 456-1097, 458-896, 458-932, 524-1128, 526-1338, 544-1061, 609-1440, 649-1440, 662-1115, 679-1440, 767-1104, 772-1258, 778-1399, 859-1440, 886-1440, 893-1111, 896-1438, 907-1412, 912-1440, 935-1393, 1005-1440, 1022-1432, 1041-1440, 1068-1326
43/7512289CB1/ 1066	1-205, 20-162, 20-276, 21-255, 30-457, 31-326, 32-323, 32-507, 33-315, 34-277, 34-314, 35-285, 36-130, 36-269, 36-271, 36-275, 36-280, 36-283, 36-289, 36-308, 36-312, 36-319, 36-322, 36-323, 36-331, 36-628, 36-637, 36-654, 38-323, 39-336, 39-654, 40-275, 40-285, 40-303, 40-315, 40-609, 41-304, 41-552, 43-352, 43-619, 44-336, 45-279, 45-512, 46-266, 47-300, 48-199, 48-566, 48-628, 49-342, 49-349, 50-261, 55-298, 55-335, 56-1058, 57-633, 58-299, 58-332, 59-260, 59-300, 59-602, 62-321, 62-648, 63-311, 63-314, 63-340, 63-364, 63-382, 63-462, 63-582, 63-657, 64-243, 64-322, 64-331, 64-659, 65-281, 65-284, 65-300, 65-309, 65-326, 65-338, 65-340, 65-345, 65-349, 65-355, 65-358, 65-372, 65-548, 65-589, 65-616, 65-659, 66-291, 66-309, 66-310, 66-327, 66-334, 66-343, 66-439, 66-659, 67-180, 67-271, 67-341, 67-342, 67-359, 67-373, 67-659, 68-155, 68-370, 69-369, 69-373, 70-301, 70-313, 70-319, 70-320, 70-337, 70-351, 70-359, 70-368, 70-573, 71-276, 71-303, 71-313, 71-422, 72-321, 72-345, 72-349, 72-580, 72-584, 73-269, 73-306, 73-314, 73-358, 73-444, 74-659, 75-321, 75-466, 76-273, 76-351, 76-362, 76-604, 77-234, 77-266, 77-275, 77-281, 77-287, 77-319, 77-320, 77-321, 77-322, 77-331, 77-349, 77-359, 77-362,

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
43 cont	77-367, 77-377, 77-387, 78-301, 78-317, 79-347, 79-362, 79-390, 80-186, 80-195, 80-280, 80-314, 80-320, 80-330, 80-335, 80-352, 80-361, 81-332, 81-345, 82-270, 82-314, 82-328, 82-329, 82-337, 82-344, 82-348, 82-351, 82-357, 82-367, 82-380, 82-448, 82-659, 83-483, 83-543, 83-630, 84-358, 87-335, 87-359, 88-287, 88-457, 89-307, 89-320, 89-373, 89-374, 90-365, 91-654, 97-344, 97-550, 99-623, 99-660, 99-685, 101-329, 101-394, 101-548, 105-389, 106-379, 106-628, 108-426, 111-378, 112-550, 116-415, 118-558, 125-330, 133-381, 133-413, 133-416, 133-435, 138-371, 139-322, 139-354, 143-379, 150-453, 151-390, 151-410, 157-402, 160-372, 160-659, 169-451, 172-506, 175-429, 176-434, 176-451, 180-445, 182-448, 185-464, 186-645, 188-342, 189-524, 190-354, 192-464, 193-322, 193-407, 193-471, 194-345, 194-447, 194-465, 194-502, 195-466, 196-415, 196-419, 196-439, 198-309, 198-423, 198-438, 198-444, 198-453, 198-659, 199-443, 199-444, 199-447, 199-448, 199-460, 199-472, 199-477, 200-479, 201-486, 201-573, 203-606, 205-427, 205-436, 205-452, 205-461, 205-462, 205-468, 206-574, 206-606, 208-624, 210-450, 214-457, 220-614, 221-468, 223-513, 233-604, 236-479, 236-505, 236-507, 236-544, 237-518,
	239-483, 240-502, 240-521, 242-385, 242-436, 242-482, 243-377, 243-470, 245-516, 247-520, 248-486, 250-508, 253-588, 254-361, 254-501, 254-515, 254-570, 256-524, 257-501, 257-512, 259-477, 260-540, 260-555, 260-571, 265-521, 265-534, 265-538, 266-450, 267-551; 271-504, 271-520, 271-529, 271-539, 271-600, 274-530, 275-488, 275-533, 277-609, 281-503, 281-541, 281-563, 281-568, 281-572, 283-511, 285-490, 285-553, 285-558, 286-525, 286-562, 291-545, 298-646, 298-657, 300-595, 302-551, 306-536, 306-610, 307-555, 309-568, 309-594, 309-654, 310-614, 312-508, 312-565, 322-508, 322-567, 322-582, 324-496, 326-543, 326-604, 327-556, 329-561, 329-576, 330-563, 330-570, 331-527, 335-584, 336-554, 336-556, 336-558, 336-582, 338-597, 341-659, 342-647, 343-602, 344-616, 344-638, 345-428, 345-540, 345-589, 345-607, 345-616, 346-644, 348-549, 348-570, 348-589, 348-618, 348-627, 350-537, 350-569, 350-582, 351-619, 351-639, 355-625, 357-563, 357-616, 358-596, 360-622, 361-630, 362-508, 362-509, 362-653, 362-659, 365-630, 366-659, 367-533, 367-592, 367-617, 367-618, 370-534, 370-659, 377-611, 377-659, 379-549, 379-649, 380-640, 382-502, 387-557, 388-619, 388-646, 388-659, 389-503,

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
43 cont	389-561, 389-630, 389-639, 389-659, 390-613, 391-616, 391-659, 397-656, 398-632, 402-659, 406-624, 406-627, 406-659, 408-580, 408-645, 408-650, 408-659, 411-638, 412-652, 413-638, 414-631, 414-642, 415-647, 417-599, 417-659, 420-659, 421-659, 422-660, 423-582, 423-641, 423-659, 426-659, 428-659, 429-659, 434-596, 434-654, 439-659, 447-659, 448-659, 456-659, 459-650, 459-659, 460-659, 464-652, 467-630, 470-657, 471-659, 472-659, 475-659, 476-623, 482-648, 486-784, 498-598, 514-659, 521-623, 521-659, 531-659, 544-659, 563-648, 585-659, 668-909, 698-962, 742-946, 754-1000, 823-1064, 880-1066, 925-1063, 927-1060, 936-1056, 946-1060, 972-1066, 973-1060, 986-1060, 987-1060
44/7512447CB1/ 4179	1-162, 52-318, 81-584, 81-612, 84-4056, 103-606, 104-532, 141-612, 141-747, 141-916, 141-932, 142-612, 395-634, 928-1182, 931-1043, 931-1189, 977-1604, 978-1604, 1508-2021, 1581-2075, 1581-2077, 1598-2074, 1721-2310, 1984-2325, 2018-2425, 2018-2614, 2023-2852, 2054-2691, 2085-2723, 2108-2580, 2144-2733, 2146-2407, 2146-2418, 2146-2477, 2163-2350, 2163-2384, 2163-2465, 2232-2935, 2234-2639, 2533-2812, 2586-2843, 2704-3149, 2829-3128, 2829-3141, 2948-3459, 2948-3509, 2953-3610, 2958-3353, 2970-3763, 3033-3264, 3033-3310, 3050-3315, 3086-3411, 3105-3706, 3119-3378, 3291-3828, 3318-4056, 3320-3590, 3376-3752, 3462-3963, 3462-4106, 3488-4179, 3515-3952, 3515-4017, 3515-4042, 3515-4065, 3697-3959, 3702-3942, 3702-3959, 3771-3896, 3773-3959, 3800-3959

Table 5

Polynucleotide SEQ ID NO:	Incyte Project ID:	Representative Library
23	7506758CB1	PGANNOT01
24	4381669CB1	BRAXTDR12
25	7503554CB1	TLYMNOT03
26	7506077CB1	OVARNOT03
27	7506765CB1	SCORNON02
28	7490048CB1	BRAIFEC01
29	2133585CB1	ENDCNOT01
30	7509063CB1	PANCTUT02
31	5496710CB1	BRAUNOR01
32	72150826CB1	COLNTUT06
33	6799476CB1	FIBRTXS07
34	7509488CB1	BRANDIT04
35	7510060CB1	CORPNOT02
36	7510226CB1	KIDEUNE02
37	7510385CB1	SINTNOR01
38	7511618CB1	ADRETUE04
39	6244135CB1	TESTNOT17
40	7506689CB1	PLACFER01
41	7510185CB1	LNOMNOF03
42	1420867CB1	BLADTUT03
43	7512289CB1	MPHGNOT03
44	7512447CB1	SINTNOR01

Table 6

Library	Vector	Library Description
ADRETUE04	PCDNA2.1	This 5 prime biased random primed library was constructed using RNA isolated from adrenal tumor tissue removed from a 52-year-old Caucasian female during a unilateral adrenalectomy. Pathology indicated a pheochromocytoma. Patient history included benign hypertension, depressive disorder, chronic sinusitis, idiopathic proctocolitis, a cataract, and urinary tract infection. Previous surgeries included a vaginal hysterectomy. Patient medications included Procardia (one dose only) and Prozac for 5 years. Family history included secondary Parkinsonism in the father; cerebrovascular disease, secondary Parkinsonism and anxiety state in the mother; and benign hypertension, atherosclerotic coronary artery disease, hyperlipidemia, and brain cancer in the sibling(s).
BLADTUT03	pINCY	Library was constructed using RNA isolated from bladder tumor tissue removed from a 58-year-old Caucasian male during a radical cystectomy, radical prostatectomy, regional lymph node excision, and urinary diversion to bowel. Pathology indicated invasive grade 3 transitional cell carcinoma. Patient history included a benign colon neoplasm. Family history included cerebrovascular disease and atherosclerotic coronary artery disease.
BRAIFEC01	pINCY	This large size-fractionated library was constructed using RNA isolated from brain tissue removed from a Caucasian male fetus who was stillborn with a hypoplastic left heart at 23 weeks' gestation.
BRANDIT04	pINCY	Library was constructed using RNA isolated from pineal gland tissue removed from a 68-year-old Caucasian female who died from congestive heart failure. Neuropathology indicated mild to moderate Alzheimer disease, atherosclerosis, and multiple infarctions. Microscopically, there were diffuse and neuritic amyloid plaques throughout the cerebral cortex. There were neurofibrillary tangles in the temporal lobes particularly the entorhinal cortex. The frontal cortex contained scattered, ballooned neurons. The amygdala contained marked gliosis, neuritic plaques and intracellular neurofibrillary tangles. The hippocampus contained neuritic and diffuse plaques, and neurofibrillary tangles. The thalamus contained diffuse and focal neuritic amyloid plaques and scattered neurofibrillary tangles. There was area of cystic cavitation with surrounding gliosis in the left globus pallidus. The pallidum contained scattered intracellular neurofibrillary tangles.
		The caudate, putamen and nucleus accumbens contained diffuse plaques. There was an area of cystic cavitation with lipid-laden macrophages in the right cerebellar hemisphere. Patient history included diabetes mellitus, rheumatoid arthritis, hyperthyroidism, amyloid heart disease, and dementia.

Table 6

Library	Vector	Library Description
BRAUNOR01	pINCY	This random primed library was constructed using RNA isolated from striatum, globus pallidus and posterior putamen tissue removed from an 81-year-old Caucasian female who died from a hemorrhage and ruptured thoracic aorta due to atherosclerosis. Pathology indicated moderate atherosclerosis involving the internal carotids, bilaterally; microscopic infarcts of the frontal cortex and hippocampus, and scattered diffuse amyloid plaques and neurofibrillary tangles, consistent with age. Grossly, the leptomeninges showed only mild thickening and hyalinization along the superior sagittal sinus. The remainder of the leptomeninges was thin and contained some congested blood vessels. Mild atrophy was found mostly in the frontal poles and lobes, and temporal lobes, bilaterally. Microscopically, there were pairs of Alzheimer type II astrocytes within the
		deep layers of the neocortex. There was increased satellitosis around neurons in the deep gray matter in the middle frontal cortex. The amygdala contained rare diffuse plaques and neurofibrillary tangles. The posterior hippocampus contained a microscopic area of cystic cavitation with hemosiderin-laden macrophages surrounded by reactive gliosis. Patient history included sepsis, cholangitis, post-operative atelectasis, pneumonia CAD, cardiomegaly due to left ventricular hypertrophy, splenomegaly, arteriolonephrosclerosis, nodular colloidal goiter, emphysema, CHF, hypothyroidism, and peripheral vascular disease.
BRAXTDR12	PCDNA2.1	This random primed library was constructed using RNA isolated from frontal neocortex tissue removed from a 55-year-old Caucasian female who died from cholangiocarcinoma. Pathology indicated mild meningeal fibrosis predominately over the convexities, scattered axonal spheroids in the white matter of the cingulate cortex and the thalamus, and a few scattered neurofibrillary tangles in the entorhinal cortex and the periaqueductal gray region. Pathology for the associated tumor tissue indicated well-differentiated cholangiocarcinoma of the liver with residual or relapsed tumor. Patient history included cholangiocarcinoma, post-operative Budd-Chiari syndrome, biliary ascites, hydrothorax, dehydration, malnutrition, oliguria and acute renal failure. Previous surgeries included cholecystectomy and resection of 85% of the liver.
COLNTUT06	pINCY	Library was constructed using RNA isolated from colon tumor tissue obtained from a 45-year-old Caucasian female during a total colectomy and total abdominal hysterectomy. Pathology indicated invasive grade 2 colonic adenocarcinoma forming a cecal mass. Patient history included benign neoplasms of the rectum and anus, multiple sclerosis and mitral valve disorder. Previous surgeries included a polypectomy. Family history included type I diabetes, cerebrovascular disease, atherosclerotic coronary artery disease, malignant skin neoplasm, hypertension, atherosclerotic coronary artery disease and malignant neoplasm of the colon.
CORPNOT02	pINCY	Library was constructed using RNA isolated from diseased corpus callosum tissue removed from the brain of a 74-year-old Caucasian male who died from Alzheimer's disease.

Table 6

Library	Vector	Library Description
ENDCNOT01	pINCY	Library was constructed using RNA isolated from endothelial cells removed from the coronary artery of a 58-year-old Hispanic male.
FIBRTXS07	pINCY	This subtracted library was constructed using 1.3 million clones from a dermal fibroblast library and was subjected to two rounds of subtraction hybridization with 2.8 million clones from an untreated dermal fibroblast tissue library. The starting library for subtraction was constructed using RNA isolated from treated dermal fibroblast tissue removed from the breast of a 31-year-old Caucasian female. The cells were treated with 9CIS retinoic acid. The hybridization probe for subtraction was derived from a similarly constructed library from RNA isolated from untreated dermal fibroblast tissue from the same donor. Subtractive hybridization conditions were based on the methodologies of Swaroop et al., NAR (1991) 19:1954 and Bonaldo, et al., Genome Research (1996) 6:791.
KIDEUNE02	pINCY	This 5' biased random primed library was constructed using RNA isolated from an untreated transformed embryonal cell line (293-EBNA) derived from kidney epithelial tissue (Invitrogen). The cells were transformed with adenovirus 5 DNA.
LNOMNOF03	PCMV-ICIS	Library was constructed using RNA isolated from mesenteric lymph node tissue removed from a 13-year-old Caucasian male who died from intracranial bleed. Serologies were negative. Previous surgeries included tonsillectomy.
MPHGNOT03	PBLUESCRIPT	Library was constructed using RNA isolated from plastic adherent mononuclear cells isolated from buffy coat units obtained from unrelated male and female donors.
OVARNOT03	PSPORT1	Library was constructed using RNA isolated from ovarian tissue removed from a 43-year-old Caucasian female during removal of the fallopian tubes and ovaries. Pathology for the associated tumor tissue indicated grade 2 mucinous cystadenocarcinoma. Patient history included mitral valve disorder, pneumonia, and viral hepatitis. Family history included atherosclerotic coronary artery disease, pancreatic cancer, stress reaction, cerebrovascular disease, breast cancer, and uterine cancer.
PANCTUT02	pINCY	Library was constructed using RNA isolated from pancreatic tumor tissue removed from a 45-year-old Caucasian female during radical pancreaticoduodenectomy. Pathology indicated a grade 4 anaplastic carcinoma. Family history included benign hypertension, hyperlipidemia and atherosclerotic coronary artery disease.
PGANNOT01	PSPORT1	Library was constructed using RNA isolated from paraganglionic tumor tissue removed from the intra-abdominal region of a 46-year-old Caucasian male during exploratory laparotomy. Pathology indicated a benign paraganglioma and was associated with a grade 2 renal cell carcinoma, clear cell type, which did not penetrate the capsule. Surgical margins were negative for tumor.

Table 6

Library	Vector	Library Description
PLACFER01	pINCY	The library was constructed using RNA isolated from placental tissue removed from a Caucasian fetus, who died after 16 weeks' gestation from fetal demise and hydrocephalus. Patient history included umbilical cord wrapped around the head (3 times) and the shoulders (1 time). Serology was positive for anti-CMV. Family history included multiple pregnancies and live births, and an abortion.
SCORNON02	PSPORT1	This normalized spinal cord library was constructed from 3.24M independent clones from the a spinal cord tissue library. RNA was isolated from the spinal cord tissue removed from a 71-year-old Caucasian male who died from respiratory arrest. Patient history included myocardial infarction, gangrene, and end stage renal disease. The normalization and hybridization conditions were adapted from Soares et al.(PNAS (1994) 91:9228).
SINTNOR01	PCDNA2.1	This random primed library was constructed using RNA isolated from small intestine tissue removed from a 31-year-old Caucasian female during Roux-en-Y gastric bypass. Patient history included clinical obesity.
TESTNOT17	pINCY	Library was constructed from testis tissue removed from a 26-year-old Caucasian male who died from head trauma due to a motor vehicle accident. Serologies were negative. Patient history included a hernia at birth, tobacco use (1 1/2 ppd), marijuana use, and daily alcohol use (beer and hard liquor).
TLYMNOT03	pINCY	Library was constructed using RNA isolated from nonactivated Th1 cells. These cells were differentiated from umbilical cord CD4 T cells with IL-12 and B7-transfected COS cells.

Table 7

Program	Description	Reference	Parameter Threshold
ABI FACTURA	A program that removes vector sequences and masks ambiguous bases in nucleic acid sequences.	Applied Biosystems, Foster City, CA.	
ABI/PARACEL FDF	A Fast Data Finder useful in comparing and annotating amino acid or nucleic acid sequences.	Applied Biosystems, Foster City, CA; Paracel Inc., Pasadena, CA.	Mismatch <50%
ABI AutoAssembler	A program that assembles nucleic acid sequences.	Applied Biosystems, Foster City, CA.	
BLAST	A Basic Local Alignment Search Tool useful in sequence similarity search for amino acid and nucleic acid sequences. BLAST includes five functions: blastp, blastn, blastx, tblastn, and tblastx.	Altschul, S.F. et al. (1990) J. Mol. Biol. 215:403-410; Altschul, S.F. et al. (1997) Nucleic Acids Res. 25:3389-3402.	ESTs: Probability value = 1.0E-8 or less; Full Length sequences: Probability value = 1.0E-10 or less
FASTA	A Pearson and Lipman algorithm that searches for similarity between a query sequence and a group of sequences of the same type. FASTA comprises as least five functions: fasta, tfasta, fastx, tfastx, and ssearch.	Pearson, W.R. and D.J. Lipman (1988) Proc. Natl. Acad Sci. USA 85:2444-2448; Pearson, W.R. (1990) Methods Enzymol. 183:63-98; and Smith, T.F. and M.S. Waterman (1981) Adv. Appl. Math. 2:482-489.	ESTs: fasta E value = 1.06E-6; Assembled ESTs: fasta Identity = 95% or greater and Match length = 200 bases or greater; fastx E value = 1.0E-8 or less; Full Length sequences: fastx score = 100 or greater
BLIMPS	A BLocks IMProved Searcher that matches a sequence against those in BLOCKS, PRINTS, DOMO, PRODOM, and PFAM databases to search for gene families, sequence homology, and structural fingerprint regions.	Henikoff, S. and J.G. Henikoff (1991) Nucleic Acids Res. 19:6565-6572; Henikoff, J.G. and S. Henikoff (1996) Methods Enzymol. 266:88-105; and Attwood, T.K. et al. (1997) J. Chem. Inf. Comput. Sci. 37:417-424.	Probability value = 1.0E-3 or less

Table 7

Program	Description	Reference	Parameter Threshold
HMMER	An algorithm for searching a query sequence against hidden Markov model (HMM)-based databases of protein family consensus sequences, such as PFAM, INCY, SMART and TIGRFAM.	Krogh, A. et al. (1994) J. Mol. Biol. 235:1501-1531; Sonnhammer, E.L.L. et al. (1988) Nucleic Acids Res. 26:320-322; Durbin, R. et al. (1998) Our World View, in a Nutshell, Cambridge Univ. Press, pp. 1-350.	PFAM, INCY, SMART or TIGRFAM hits: Probability value = 1.0E-3 or less; Signal peptide hits: Score = 0 or greater
ProfileScan	An algorithm that searches for structural and sequence motifs in protein sequences that match sequence patterns defined in Prosite.	Gribskov, M. et al. (1988) CABIOS 4:61-66; Gribskov, M. et al. (1989) Methods Enzymol. 183:146-159; Bairoch, A. et al. (1997) Nucleic Acids Res. 25:217-221.	Normalized quality score ≥ GCG specified "HIGH" value for that particular Prosite motif. Generally, score = 1.4-2.1.
Phred	A base-calling algorithm that examines automated sequencer traces with high sensitivity and probability.	Ewing, B. et al. (1998) Genome Res. 8:175-185; Ewing, B. and P. Green (1998) Genome Res. 8:186-194.	
Phrap	A Phils Revised Assembly Program including SWAT and CrossMatch, programs based on efficient implementation of the Smith-Waterman algorithm, useful in searching sequence homology and assembling DNA sequences.	Smith, T.F. and M.S. Waterman (1981) Adv. Appl. Math. 2:482-489; Smith, T.F. and M.S. Waterman (1981) J. Mol. Biol. 147:195-197; and Green, P., University of Washington, Seattle, WA.	Score = 120 or greater; Match length = 56 or greater
Consed	A graphical tool for viewing and editing Phrap assemblies.	Gordon, D. et al. (1998) Genome Res. 8:195-202.	
SPScan	A weight matrix analysis program that scans protein sequences for the presence of secretory signal peptides.	Nielson, H. et al. (1997) Protein Engineering 10:1-6; Claverie, J.M. and S. Audic (1997) CABIOS 12:431-439.	Score = 3.5 or greater
TMAP	A program that uses weight matrices to delineate transmembrane segments on protein sequences and determine orientation.	Persson, B. and P. Argos (1994) J. Mol. Biol. 237:182-192; Persson, B. and P. Argos (1996) Protein Sci. 5:363-371.	

Table 7

Program	Description	Reference	Parameter Threshold
TMHMMER	A program that uses a hidden Markov model (HMM) to delineate transmembrane segments on protein sequences and determine orientation.	Sonnhammer, E.L. et al. (1998) Proc. Sixth Intl. Conf. On Intelligent Systems for Mol. Biol., Glasgow et al., eds., The Am. Assoc. for Artificial Intelligence (AAAI) Press, Menlo Park, CA, and MIT Press, Cambridge, MA, pp. 175-182.	
Motifs	A program that searches amino acid sequences for patterns that matched those defined in Prosite.	Bairoch, A. et al. (1997) Nucleic Acids Res. 25:217-221; Wisconsin Package Program Manual, version 9, page M51-59, Genetics Computer Group, Madison, WI.	

What is claimed is:

1. An isolated polypeptide selected from the group consisting of:

a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-22,

b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-2, SEQ ID NO:5-6, SEQ ID NO:10, SEQ ID NO:15, and SEQ ID NO:17-19,

c) a polypeptide comprising a naturally occurring amino acid sequence at least 95% identical to the amino acid sequence of SEQ ID NO:3,

d) a polypeptide comprising a naturally occurring amino acid sequence at least 96% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:9, and SEQ ID NO:14,

e) a polypeptide comprising a naturally occurring amino acid sequence at least 97% identical to the amino acid sequence of SEQ ID NO:7,

f) a polypeptide comprising a naturally occurring amino acid sequence at least 99% identical to an amino acid sequence selected from the group consisting of of SEQ ID NO:13, and SEQ ID NO:20,

g) a polypeptide consisting essentially of a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of of SEQ ID NO:4, SEQ ID NO:8, SEQ ID NO:11-12, SEQ ID NO:16, and SEQ ID NO:21-22,

h) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-22, and

i) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-22.

2. An isolated polypeptide of claim 1 comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-22.

3. An isolated polynucleotide encoding a polypeptide of claim 1.

4. An isolated polynucleotide encoding a polypeptide of claim 2.

5. An isolated polynucleotide of claim 4 comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:23-44.

6. A recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide of claim 3.

7. A cell transformed with a recombinant polynucleotide of claim 6.

8. A transgenic organism comprising a recombinant polynucleotide of claim 6.

9. A method of producing a polypeptide of claim 1, the method comprising:

a) culturing a cell under conditions suitable for expression of the polypeptide, wherein said cell is transformed with a recombinant polynucleotide, and said recombinant polynucleotide comprises a promoter sequence operably linked to a polynucleotide encoding the polypeptide of claim 1, and

b) recovering the polypeptide so expressed.

10. A method of claim 9, wherein the polypeptide comprises an amino acid sequence selected from the group consisting of SEQ ID NO:1-22.

11. An isolated antibody which specifically binds to a polypeptide of claim 1.

12. An isolated polynucleotide selected from the group consisting of:

a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:23-44,

b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:23-28, SEQ ID NO:31-34, SEQ ID NO:36-39, and SEQ ID NO:41,

c) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 93% identical to the polynucleotide sequence of SEQ ID NO:40,

d) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 94% identical to the polynucleotide sequence of SEQ ID NO:30,

e) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 96% identical to the polynucleotide sequence of SEQ ID NO:44,

f) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 97% identical to the polynucleotide sequence of SEQ ID NO:35,

g) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 99% identical to the polynucleotide sequence of SEQ ID NO:43,

h) a polynucleotide consisting essentially of a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:29, and SEQ ID NO:42

i) a polynucleotide complementary to a polynucleotide of a),

j) a polynucleotide complementary to a polynucleotide of b),

k) a polynucleotide complementary to a polynucleotide of c),

l) a polynucleotide complementary to a polynucleotide of d),

m) a polynucleotide complementary to a polynucleotide of e),

n) a polynucleotide complementary to a polynucleotide of f),

o) a polynucleotide complementary to a polynucleotide of g),

p) a polynucleotide complementary to a polynucleotide of h), and

q) an RNA equivalent of a)-p).

13. An isolated polynucleotide comprising at least 60 contiguous nucleotides of a polynucleotide of claim 12.

14. A method of detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide of claim 12, the method comprising:

a) hybridizing the sample with a probe comprising at least 20 contiguous nucleotides comprising a sequence complementary to said target polynucleotide in the sample, and which probe specifically hybridizes to said target polynucleotide, under conditions whereby a hybridization complex is formed between said probe and said target polynucleotide or fragments thereof, and

b) detecting the presence or absence of said hybridization complex, and, optionally, if

present, the amount thereof.

15. A method of claim 14, wherein the probe comprises at least 60 contiguous nucleotides.

16. A method of detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide of claim 12, the method comprising:

a) amplifying said target polynucleotide or fragment thereof using polymerase chain reaction amplification, and

b) detecting the presence or absence of said amplified target polynucleotide or fragment thereof, and, optionally, if present, the amount thereof.

17. A composition comprising a polypeptide of claim 1 and a pharmaceutically acceptable excipient.

18. A composition of claim 17, wherein the polypeptide comprises an amino acid sequence selected from the group consisting of SEQ ID NO:1-22.

19. A method for treating a disease or condition associated with decreased expression of functional NTRAN, comprising administering to a patient in need of such treatment the composition of claim 17.

20. A method of screening a compound for effectiveness as an agonist of a polypeptide of claim 1, the method comprising:

a) exposing a sample comprising a polypeptide of claim 1 to a compound, and

b) detecting agonist activity in the sample.

21. A composition comprising an agonist compound identified by a method of claim 20 and a pharmaceutically acceptable excipient.

22. A method for treating a disease or condition associated with decreased expression of functional NTRAN, comprising administering to a patient in need of such treatment a composition of claim 21.

23. A method of screening a compound for effectiveness as an antagonist of a polypeptide of claim 1, the method comprising:

a) exposing a sample comprising a polypeptide of claim 1 to a compound, and

b) detecting antagonist activity in the sample.

24. A composition comprising an antagonist compound identified by a method of claim 23 and a pharmaceutically acceptable excipient.

25. A method for treating a disease or condition associated with overexpression of functional NTRAN, comprising administering to a patient in need of such treatment a composition of claim 24.

26. A method of screening for a compound that specifically binds to the polypeptide of claim 1, the method comprising:

a) combining the polypeptide of claim 1 with at least one test compound under suitable conditions, and

b) detecting binding of the polypeptide of claim 1 to the test compound, thereby identifying a compound that specifically binds to the polypeptide of claim 1.

27. A method of screening for a compound that modulates the activity of the polypeptide of claim 1, the method comprising:

a) combining the polypeptide of claim 1 with at least one test compound under conditions permissive for the activity of the polypeptide of claim 1,

b) assessing the activity of the polypeptide of claim 1 in the presence of the test compound, and

c) comparing the activity of the polypeptide of claim 1 in the presence of the test compound with the activity of the polypeptide of claim 1 in the absence of the test compound, wherein a change in the activity of the polypeptide of claim 1 in the presence of the test compound is indicative of a compound that modulates the activity of the polypeptide of claim 1.

28. A method of screening a compound for effectiveness in altering expression of a target polynucleotide, wherein said target polynucleotide comprises a sequence of claim 5, the method

comprising:

a) exposing a sample comprising the target polynucleotide to a compound, under conditions suitable for the expression of the target polynucleotide,

b) detecting altered expression of the target polynucleotide, and

c) comparing the expression of the target polynucleotide in the presence of varying amounts of the compound and in the absence of the compound.

29. A method of assessing toxicity of a test compound, the method comprising:

a) treating a biological sample containing nucleic acids with the test compound,

b) hybridizing the nucleic acids of the treated biological sample with a probe comprising at least 20 contiguous nucleotides of a polynucleotide of claim 12 under conditions whereby a specific hybridization complex is formed between said probe and a target polynucleotide in the biological sample, said target polynucleotide comprising a polynucleotide sequence of a polynucleotide of claim 12 or fragment thereof,

c) quantifying the amount of hybridization complex, and

d) comparing the amount of hybridization complex in the treated biological sample with the amount of hybridization complex in an untreated biological sample, wherein a difference in the amount of hybridization complex in the treated biological sample is indicative of toxicity of the test compound.

30. A method for a diagnostic test for a condition or disease associated with the expression of NTRAN in a biological sample, the method comprising:

a) combining the biological sample with an antibody of claim 11, under conditions suitable for the antibody to bind the polypeptide and form an antibody:polypeptide complex, and

b) detecting the complex, wherein the presence of the complex correlates with the presence of the polypeptide in the biological sample.

31. The antibody of claim 11, wherein the antibody is:

a) a chimeric antibody,

b) a single chain antibody,

c) a Fab fragment,

d) a $F(ab')_2$ fragment, or

e) a humanized antibody.

32. A composition comprising an antibody of claim 11 and an acceptable excipient.

33. A method of diagnosing a condition or disease associated with the expression of NTRAN in a subject, comprising administering to said subject an effective amount of the composition of claim 32.

34. A composition of claim 32, further comprising a label.

35. A method of diagnosing a condition or disease associated with the expression of NTRAN in a subject, comprising administering to said subject an effective amount of the composition of claim 34.

36. A method of preparing a polyclonal antibody with the specificity of the antibody of claim 11, the method comprising:

a) immunizing an animal with a polypeptide consisting of an amino acid sequence selected from the group consisting of SEQ ID NO:1-22, or an immunogenic fragment thereof, under conditions to elicit an antibody response,

b) isolating antibodies from the animal, and

c) screening the isolated antibodies with the polypeptide, thereby identifying a polyclonal antibody which specifically binds to a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-22.

37. A polyclonal antibody produced by a method of claim 36.

38. A composition comprising the polyclonal antibody of claim 37 and a suitable carrier.

39. A method of making a monoclonal antibody with the specificity of the antibody of claim 11, the method comprising:

a) immunizing an animal with a polypeptide consisting of an amino acid sequence selected from the group consisting of SEQ ID NO:1-22, or an immunogenic fragment thereof, under conditions to elicit an antibody response,

b) isolating antibody producing cells from the animal,

c) fusing the antibody producing cells with immortalized cells to form monoclonal antibody-producing hybridoma cells,

d) culturing the hybridoma cells, and

e) isolating from the culture monoclonal antibody which specifically binds to a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-22.

40. A monoclonal antibody produced by a method of claim 39.

41. A composition comprising the monoclonal antibody of claim 40 and a suitable carrier.

42. The antibody of claim 11, wherein the antibody is produced by screening a Fab expression library.

43. The antibody of claim 11, wherein the antibody is produced by screening a recombinant immunoglobulin library.

44. A method of detecting a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-22 in a sample, the method comprising:

a) incubating the antibody of claim 11 with the sample under conditions to allow specific binding of the antibody and the polypeptide, and

b) detecting specific binding, wherein specific binding indicates the presence of a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-22 in the sample.

45. A method of purifying a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-22 from a sample, the method comprising:

a) incubating the antibody of claim 11 with the sample under conditions to allow specific binding of the antibody and the polypeptide, and

b) separating the antibody from the sample and obtaining the purified polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-22.

46. A microarray wherein at least one element of the microarray is a polynucleotide of claim 13.

47. A method of generating an expression profile of a sample which contains polynucleotides, the method comprising:

a) labeling the polynucleotides of the sample,

b) contacting the elements of the microarray of claim 46 with the labeled polynucleotides of the sample under conditions suitable for the formation of a hybridization complex, and

c) quantifying the expression of the polynucleotides in the sample.

48. An array comprising different nucleotide molecules affixed in distinct physical locations on a solid substrate, wherein at least one of said nucleotide molecules comprises a first oligonucleotide or polynucleotide sequence specifically hybridizable with at least 30 contiguous nucleotides of a target polynucleotide, and wherein said target polynucleotide is a polynucleotide of claim 12.

49. An array of claim 48, wherein said first oligonucleotide or polynucleotide sequence is completely complementary to at least 30 contiguous nucleotides of said target polynucleotide.

50. An array of claim 48, wherein said first oligonucleotide or polynucleotide sequence is completely complementary to at least 60 contiguous nucleotides of said target polynucleotide.

51. An array of claim 48, wherein said first oligonucleotide or polynucleotide sequence is completely complementary to said target polynucleotide.

52. An array of claim 48, which is a microarray.

53. An array of claim 48, further comprising said target polynucleotide hybridized to a nucleotide molecule comprising said first oligonucleotide or polynucleotide sequence.

54. An array of claim 48, wherein a linker joins at least one of said nucleotide molecules to said solid substrate.

55. An array of claim 48, wherein each distinct physical location on the substrate contains multiple nucleotide molecules, and the multiple nucleotide molecules at any single distinct physical location have the same sequence, and each distinct physical location on the substrate contains nucleotide molecules having a sequence which differs from the sequence of nucleotide molecules at another distinct physical location on the substrate.

56. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:1.

57. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:2.

58. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:3.

59. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:4.

60. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:5.

61. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:6.

62. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:7.

63. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:8.

64. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:9.

65. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:10.

66. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:11.

67. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:12.

68. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:13.

69. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:14.

70. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:15.

71. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:16.

72. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:17.

73. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:18.

74. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:19.

75. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:20.

76. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:21.

77. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:22.

78. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:23.

79. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:24.

80. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:25.

81. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:26.

82. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:27.

83. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:28.

84. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:29.

85. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:30.

86. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:31.

87. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:32.

88. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:33.

89. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:34.

90. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:35.

91. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:36.

92. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:37.

93. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:38.

94. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:39.

95. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:40.

96. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:41.

97. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:42.

98. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:43.

99. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:44.

<110> INCYTE GENOMICS, INC.
BAUGHN, Mariah R.
BHATIA, Umesh G.
BLAKE, Julie J.
BURRILL, John D.
ELLIOT, Vicki S.
EMERLING, Brooke M.
FORSYTHE, Ian J.
GIETZEN, Kimberly J.
GORVAD, Ann E.
GRIFFIN, Jennifer A.
HAFALIA, April J.A.
HO, Anne
JACKSON, Alan A.
JIANG, Xin
KABLE, Amy E.
KEARNEY Liam
KHARE, Reena
LEE, Ernestine A.
LEE, Sally
LU, Dyung Aina M.
MARQUIS, Joseph P.
LEHR-MASON, Patricia M.
RAMKUMAR, Jayalaxmi
RICHARDSON Thomas W.
SPRAGUE, William W.
TRAN, Uyen K.
CHAWLA, Narinder K.
WARREN, Bridget A.
YUE, Henry
ZHENG, Wenjin

<120> NEUROTRANSMISSION-ASSOCIATED PROTEINS

<130> PF-1331 PCT

<140> To Be Assigned
<141> Herewith

<150> US 60/340,798
<151> 2001-12-14

<150> US 60/365,645
<151> 2002-03-18

<150> US 60/367,662
<151> 2002-03-25

<150> US 60/379,887
<151> 2002-05-10

<150> US 60/384,639
<151> 2002-05-31

<160> 44
<170> PERL Program

<210> 1
<211> 151
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7506758CD1

<400> 1
Met Ala Gln Glu Ile Asp Leu Ser Ala Leu Lys Glu Leu Glu Arg
1 5 10 15
Glu Ala Ile Leu Gln Val Leu Tyr Arg Asp Gln Ala Val Gln Asn
20 25 30
Thr Glu Glu Glu Arg Thr Arg Lys Leu Lys Thr His Leu Gln His
35 40 45
Leu Arg Trp Lys Gly Ala Lys Asn Thr Asp Trp Glu His Lys Glu
50 55 60
Lys Cys Cys Ala Arg Cys Gln Gln Val Leu Gly Phe Leu Leu His
65 70 75
Arg Gly Ala Val Cys Arg Gly Cys Ser His Arg Val Cys Ala Gln
80 85 90
Cys Arg Val Phe Leu Arg Gly Thr His Ala Trp Lys Cys Thr Val
95 100 105
Cys Phe Glu Asp Arg Asn Val Lys Ile Lys Thr Gly Glu Trp Phe
110 115 120
Tyr Glu Glu Arg Ala Lys Lys Phe Pro Thr Gly Glu Gly Gln Cys
125 130 135
Thr Arg Tyr Ser Glu Thr Ser Gln Ser Arg Gly Ser Gln Met Leu
140 145 150
Tyr

<210> 2
<211> 2507
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 4381669CD1

<400> 2
Met Asp Ser Val Phe Ser Ile Tyr Phe Ser Met His Phe Gln Met
1 5 10 15
Arg Ile Leu Gly His Asn Ala Tyr Ala Leu Gln Gln Ala Ala Phe
20 25 30
Val Cys Leu Gln Leu Tyr Ser Arg Pro Asn Arg Leu Leu Ser Gly
35 40 45
Val Pro Leu Tyr Lys Gly Asp Val Pro Thr Gln Asp Lys Ser Ala
50 55 60
Val Thr Ser Leu Leu Asp Gly Leu Asn Gln Ala Phe Glu Glu Val
65 70 75
Ser Ser Gln Ser Gly Gly Ala Lys Arg Gln Ser Ile His Leu Glu
80 85 90
Gln Lys Leu Tyr Asp Gly Val Ser Ala Thr Ser Thr Trp Leu Asp

```
                95                  100                 105
Asp Val Glu Glu Arg Leu Phe Val Ala Thr Ala Leu Leu Pro Glu
            110                 115                 120
Glu Thr Glu Thr Cys Leu Phe Asn Gln Glu Ile Leu Ala Lys Asp
            125                 130                 135
Ile Lys Glu Met Ser Glu Glu Met Asp Lys Asn Lys Asn Leu Phe
            140                 145                 150
Ser Gln Ala Phe Pro Glu Asn Gly Asp Asn Arg Asp Val Ile Glu
            155                 160                 165
Asp Thr Leu Gly Cys Leu Leu Gly Arg Leu Ser Leu Leu Asp Ser
            170                 175                 180
Val Val Asn Gln Arg Cys His Gln Met Lys Glu Arg Leu Gln Gln
            185                 190                 195
Ile Leu Asn Phe Gln Asn Asp Leu Lys Val Leu Phe Thr Ser Leu
            200                 205                 210
Ala Asp Asn Lys Tyr Ile Ile Leu Gln Lys Leu Ala Asn Val Phe
            215                 220                 225
Glu Gln Pro Val Ala Glu Gln Ile Glu Ala Ile Gln Gln Ala Glu
            230                 235                 240
Asp Gly Leu Lys Glu Phe Asp Ala Gly Ile Ile Glu Leu Lys Arg
            245                 250                 255
Arg Gly Asp Lys Leu Gln Val Glu Gln Pro Ser Met Gln Glu Leu
            260                 265                 270
Ser Lys Leu Gln Asp Met Tyr Asp Glu Leu Met Met Ile Ile Gly
            275                 280                 285
Ser Arg Arg Ser Gly Leu Asn Gln Asn Leu Thr Leu Lys Ser Gln
            290                 295                 300
Tyr Glu Arg Ala Leu Gln Asp Leu Ala Asp Leu Leu Glu Thr Gly
            305                 310                 315
Gln Glu Lys Met Ala Gly Asp Gln Lys Ile Ile Val Ser Ser Lys
            320                 325                 330
Glu Glu Ile Gln Gln Leu Leu Asp Lys His Lys Glu Tyr Phe Gln
            335                 340                 345
Gly Leu Glu Ser His Met Ile Leu Thr Glu Thr Leu Phe Arg Lys
            350                 355                 360
Ile Ile Ser Phe Ala Val Gln Lys Glu Thr Gln Phe His Thr Glu
            365                 370                 375
Leu Met Ala Gln Ala Ser Ala Val Leu Lys Arg Ala His Lys Arg
            380                 385                 390
Gly Val Glu Leu Glu Tyr Ile Leu Glu Thr Trp Ser His Leu Asp
            395                 400                 405
Glu Asp Gln Gln Glu Leu Ser Arg Gln Leu Glu Val Val Glu Ser
            410                 415                 420
Ser Ile Pro Ser Val Gly Leu Val Glu Glu Asn Glu Asp Arg Leu
            425                 430                 435
Ile Asp Arg Ile Thr Leu Tyr Gln His Leu Lys Ser Ser Leu Asn
            440                 445                 450
Glu Tyr Gln Pro Lys Leu Tyr Gln Val Leu Asp Asp Gly Lys Arg
            455                 460                 465
Leu Leu Ile Ser Ile Ser Cys Ser Asp Leu Glu Ser Gln Leu Asn
            470                 475                 480
Gln Leu Gly Glu Cys Trp Leu Ser Asn Thr Asn Lys Met Ser Lys
            485                 490                 495
Glu Leu His Arg Leu Glu Thr Ile Leu Lys His Trp Thr Arg Tyr
            500                 505                 510
Gln Ser Glu Ser Ala Asp Leu Ile His Trp Leu Gln Ser Ala Lys
```

```
                515                 520                 525
Asp Arg Leu Glu Phe Trp Thr Gln Gln Ser Val Thr Val Pro Gln
                530                 535                 540
Glu Leu Glu Met Val Arg Asp His Leu Asn Ala Phe Leu Glu Phe
                545                 550                 555
Ser Lys Glu Val Asp Ala Gln Ser Ser Leu Lys Ser Ser Val Leu
                560                 565                 570
Ser Thr Gly Asn Gln Leu Leu Arg Leu Lys Lys Val Asp Thr Ala
                575                 580                 585
Thr Leu Arg Ser Glu Leu Ser Arg Ile Asp Ser Gln Trp Thr Asp
                590                 595                 600
Leu Leu Thr Asn Ile Pro Ala Val Gln Glu Lys Leu His Gln Leu
                605                 610                 615
Gln Met Asp Lys Leu Pro Ser Arg His Ala Ile Ser Glu Val Met
                620                 625                 630
Ser Trp Ile Ser Leu Met Glu Asn Val Ile Gln Lys Asp Glu Asp
                635                 640                 645
Asn Ile Lys Asn Ser Ile Gly Tyr Lys Ala Ile His Glu Tyr Leu
                650                 655                 660
Gln Lys Tyr Lys Gly Phe Lys Ile Asp Ile Asn Cys Lys Gln Leu
                665                 670                 675
Thr Val Asp Phe Val Asn Gln Ser Val Leu Gln Ile Ser Ser Gln
                680                 685                 690
Asp Val Glu Ser Lys Arg Ser Asp Lys Thr Asp Phe Ala Glu Gln
                695                 700                 705
Leu Gly Ala Met Asn Lys Ser Trp Gln Ile Leu Gln Gly Leu Val
                710                 715                 720
Thr Glu Lys Ile Gln Leu Leu Glu Gly Leu Leu Glu Ser Trp Ser
                725                 730                 735
Glu Tyr Glu Asn Asn Val Gln Cys Leu Lys Thr Trp Phe Glu Thr
                740                 745                 750
Gln Glu Lys Arg Leu Lys Gln Gln His Arg Ile Gly Asp Gln Ala
                755                 760                 765
Ser Val Gln Asn Ala Leu Lys Asp Cys Gln Asp Leu Glu Asp Leu
                770                 775                 780
Ile Lys Ala Lys Glu Lys Glu Val Glu Lys Ile Glu Gln Asn Gly
                785                 790                 795
Leu Ala Leu Ile Gln Asn Lys Lys Glu Asp Val Ser Ser Ile Val
                800                 805                 810
Met Ser Thr Leu Arg Glu Leu Gly Gln Thr Trp Ala Asn Leu Asp
                815                 820                 825
His Met Val Gly Gln Leu Lys Ile Leu Leu Lys Ser Val Leu Asp
                830                 835                 840
Gln Trp Ser Ser His Lys Val Ala Phe Asp Lys Ile Asn Ser Tyr
                845                 850                 855
Leu Met Glu Ala Arg Tyr Ser Leu Ser Arg Phe Arg Leu Leu Thr
                860                 865                 870
Gly Ser Leu Glu Ala Val Gln Val Gln Val Asp Asn Leu Gln Asn
                875                 880                 885
Leu Gln Asp Asp Leu Glu Lys Gln Glu Arg Ser Leu Gln Lys Phe
                890                 895                 900
Gly Ser Ile Thr Asn Gln Leu Leu Lys Glu Cys His Pro Pro Val
                905                 910                 915
Thr Glu Thr Leu Thr Asn Thr Leu Lys Glu Val Asn Met Arg Trp
                920                 925                 930
Asn Asn Leu Leu Glu Glu Ile Ala Glu Gln Leu Gln Ser Ser Lys
```

```
                935                 940                 945
Ala Leu Leu Gln Leu Trp Gln Arg Tyr Lys Asp Tyr Ser Lys Gln
                950                 955                 960
Cys Ala Ser Thr Val Gln Gln Gln Glu Asp Arg Thr Asn Glu Leu
                965                 970                 975
Leu Lys Ala Ala Thr Asn Lys Asp Ile Ala Asp Asp Glu Val Ala
                980                 985                 990
Thr Trp Ile Gln Asp Cys Asn Asp Leu Leu Lys Gly Leu Gly Thr
                995                1000                1005
Val Lys Asp Ser Leu Phe Phe Leu His Glu Leu Gly Glu Gln Leu
               1010                1015                1020
Lys Gln Gln Val Asp Ala Ser Ala Ala Ser Ala Ile Gln Ser Asp
               1025                1030                1035
Gln Leu Ser Leu Ser Gln His Leu Cys Ala Leu Glu Gln Ala Leu
               1040                1045                1050
Cys Lys Gln Gln Thr Ser Leu Gln Ala Gly Val Leu Asp Tyr Glu
               1055                1060                1065
Thr Phe Ala Lys Ser Leu Glu Ala Leu Glu Ala Trp Ile Val Glu
               1070                1075                1080
Ala Glu Glu Ile Leu Gln Gly Gln Asp Pro Ser His Ser Ser Asp
               1085                1090                1095
Leu Ser Thr Ile Gln Glu Arg Met Glu Glu Leu Lys Gly Gln Met
               1100                1105                1110
Leu Lys Phe Ser Ser Met Ala Pro Asp Leu Asp Arg Leu Asn Glu
               1115                1120                1125
Leu Gly Tyr Arg Leu Pro Leu Asn Asp Lys Glu Ile Lys Arg Met
               1130                1135                1140
Gln Asn Leu Asn Arg His Trp Ser Leu Ile Ser Ser Gln Thr Thr
               1145                1150                1155
Glu Arg Phe Ser Lys Leu Gln Ser Phe Leu Leu Gln His Gln Thr
               1160                1165                1170
Phe Leu Glu Lys Cys Glu Thr Trp Met Glu Phe Leu Val Gln Thr
               1175                1180                1185
Glu Gln Lys Leu Ala Val Glu Ile Ser Gly Asn Tyr Gln His Leu
               1190                1195                1200
Leu Glu Gln Gln Arg Ala His Glu Val Leu Phe Gln Ala Glu Met
               1205                1210                1215
Phe Ser Arg Gln Gln Ile Leu His Ser Ile Ile Ile Asp Gly Gln
               1220                1225                1230
Arg Leu Leu Glu Gln Gly Gln Val Asp Asp Arg Asp Glu Phe Asn
               1235                1240                1245
Leu Lys Leu Thr Leu Leu Ser Asn Gln Trp Gln Gly Val Ile Arg
               1250                1255                1260
Arg Ala Gln Gln Arg Arg Gly Ile Ile Asp Ser Gln Ile Arg Gln
               1265                1270                1275
Trp Gln Arg Tyr Arg Glu Met Ala Glu Lys Leu Arg Lys Trp Leu
               1280                1285                1290
Val Glu Val Ser Tyr Leu Pro Met Ser Gly Leu Gly Ser Val Pro
               1295                1300                1305
Ile Pro Leu Gln Gln Ala Arg Thr Leu Phe Asp Glu Val Gln Phe
               1310                1315                1320
Lys Glu Lys Val Phe Leu Arg Gln Gln Gly Ser Tyr Ile Leu Thr
               1325                1330                1335
Val Glu Ala Gly Lys Gln Leu Leu Leu Ser Ala Asp Ser Gly Ala
               1340                1345                1350
Glu Ala Ala Leu Gln Ala Glu Leu Ala Glu Ile Gln Glu Lys Trp
```

```
                1355                1360                1365
Lys Ser Ala Ser Met Arg Leu Glu Glu Gln Lys Lys Lys Leu Ala
                1370                1375                1380
Phe Leu Leu Lys Asp Trp Glu Lys Cys Glu Lys Gly Ile Ala Asp
                1385                1390                1395
Ser Leu Glu Lys Leu Arg Thr Phe Lys Lys Lys Leu Ser Gln Ser
                1400                1405                1410
Leu Pro Asp His His Glu Glu Leu His Ala Glu Gln Met Arg Cys
                1415                1420                1425
Lys Glu Leu Glu Asn Ala Val Gly Ser Trp Thr Asp Asp Leu Thr
                1430                1435                1440
Gln Leu Ser Leu Leu Lys Asp Thr Leu Ser Ala Tyr Ile Ser Ala
                1445                1450                1455
Asp Asp Ile Ser Ile Leu Asn Glu Arg Val Glu Leu Leu Gln Arg
                1460                1465                1470
Gln Trp Glu Glu Leu Cys His Gln Leu Ser Leu Arg Arg Gln Gln
                1475                1480                1485
Ile Gly Glu Arg Leu Asn Glu Trp Ala Val Phe Ser Glu Lys Asn
                1490                1495                1500
Lys Glu Leu Cys Glu Trp Leu Thr Gln Met Glu Ser Lys Val Ser
                1505                1510                1515
Gln Asn Gly Asp Ile Leu Ile Glu Glu Met Ile Glu Lys Leu Lys
                1520                1525                1530
Lys Asp Tyr Gln Glu Glu Ile Ala Ile Ala Gln Glu Asn Lys Ile
                1535                1540                1545
Gln Leu Gln Gln Met Gly Glu Arg Leu Ala Lys Ala Ser His Glu
                1550                1555                1560
Ser Lys Ala Ser Glu Ile Glu Tyr Lys Leu Gly Lys Val Asn Asp
                1565                1570                1575
Arg Trp Gln His Leu Leu Asp Leu Ile Ala Ala Arg Val Lys Lys
                1580                1585                1590
Leu Lys Glu Thr Leu Val Ala Val Gln Gln Leu Asp Lys Asn Met
                1595                1600                1605
Ser Ser Leu Arg Thr Trp Leu Ala His Ile Glu Ser Glu Leu Ala
                1610                1615                1620
Lys Pro Ile Val Tyr Asp Ser Cys Asn Ser Glu Glu Ile Gln Arg
                1625                1630                1635
Lys Leu Asn Glu Gln Gln Glu Leu Gln Arg Asp Ile Glu Lys His
                1640                1645                1650
Ser Thr Gly Val Ala Ser Val Leu Asn Leu Cys Glu Val Leu Leu
                1655                1660                1665
His Asp Cys Asp Ala Cys Ala Thr Asp Ala Glu Cys Asp Ser Ile
                1670                1675                1680
Gln Gln Ala Thr Arg Asn Leu Asp Arg Arg Trp Arg Asn Ile Cys
                1685                1690                1695
Ala Met Ser Met Glu Arg Arg Leu Lys Ile Glu Glu Thr Trp Arg
                1700                1705                1710
Leu Trp Gln Lys Phe Leu Asp Asp Tyr Ser Arg Phe Glu Asp Trp
                1715                1720                1725
Leu Lys Ser Ser Glu Arg Thr Ala Ala Phe Pro Ser Ser Ser Gly
                1730                1735                1740
Val Ile Tyr Thr Val Ala Lys Glu Glu Leu Lys Lys Phe Glu Ala
                1745                1750                1755
Phe Gln Arg Gln Val His Glu Cys Leu Thr Gln Leu Glu Leu Ile
                1760                1765                1770
Asn Lys Gln Tyr Arg Arg Leu Ala Arg Glu Asn Arg Thr Asp Ser
```

```
                1775                1780                1785
Ala Cys Ser Leu Lys Gln Met Val His Glu Gly Asn Gln Arg Trp
                1790                1795                1800
Asp Asn Leu Gln Lys Arg Val Thr Ser Ile Leu Arg Arg Leu Lys
                1805                1810                1815
His Phe Ile Gly Gln Arg Glu Glu Phe Glu Thr Ala Arg Asp Ser
                1820                1825                1830
Ile Leu Val Trp Leu Thr Glu Met Asp Leu Gln Leu Thr Asn Ile
                1835                1840                1845
Glu His Phe Ser Glu Cys Asp Val Gln Ala Lys Ile Lys Gln Leu
                1850                1855                1860
Lys Ala Phe Gln Gln Glu Ile Ser Leu Asn His Asn Lys Ile Glu
                1865                1870                1875
Gln Ile Ile Ala Gln Gly Glu Gln Leu Ile Glu Lys Ser Glu Pro
                1880                1885                1890
Leu Asp Ala Ala Ile Ile Glu Glu Glu Leu Asp Glu Leu Arg Arg
                1895                1900                1905
Tyr Cys Gln Glu Val Phe Gly Arg Val Glu Arg Tyr His Lys Lys
                1910                1915                1920
Leu Ile Arg Leu Pro Val Tyr Asp Glu His Asp Leu Ser Asp Arg
                1925                1930                1935
Glu Leu Glu Leu Glu Asp Ser Ala Ala Leu Ser Asp Leu His Trp
                1940                1945                1950
His Asp Arg Ser Ala Asp Ser Leu Leu Ser Pro Gln Pro Ser Ser
                1955                1960                1965
Asn Leu Ser Leu Ser Leu Ala Gln Pro Leu Arg Ser Glu Arg Ser
                1970                1975                1980
Gly Arg Asp Thr Pro Ala Ser Val Asp Ser Ile Pro Leu Glu Trp
                1985                1990                1995
Asp His Asp Tyr Asp Leu Ser Arg Asp Leu Glu Ser Ala Met Ser
                2000                2005                2010
Arg Ala Leu Pro Ser Glu Asp Glu Glu Gly Gln Asp Asp Lys Asp
                2015                2020                2025
Phe Tyr Leu Arg Gly Ala Val Gly Leu Ser Gly Asp His Ser Ala
                2030                2035                2040
Leu Glu Ser Gln Ile Arg Gln Leu Gly Lys Ala Leu Asp Asp Ser
                2045                2050                2055
Arg Phe Gln Ile Gln Gln Thr Glu Asn Ile Ile Arg Ser Lys Thr
                2060                2065                2070
Pro Thr Gly Pro Glu Leu Asp Thr Ser Tyr Lys Gly Tyr Met Lys
                2075                2080                2085
Leu Leu Gly Glu Cys Ser Ser Ser Ile Asp Ser Val Lys Arg Leu
                2090                2095                2100
Glu His Lys Leu Lys Glu Glu Glu Glu Ser Leu Pro Gly Phe Val
                2105                2110                2115
Asn Leu His Ser Thr Glu Thr Gln Thr Ala Gly Val Ile Asp Arg
                2120                2125                2130
Trp Glu Leu Leu Gln Ala Gln Ala Leu Ser Lys Glu Leu Arg Met
                2135                2140                2145
Lys Gln Asn Leu Gln Lys Trp Gln Gln Phe Asn Ser Asp Leu Asn
                2150                2155                2160
Ser Ile Trp Ala Trp Leu Gly Asp Thr Glu Glu Glu Leu Glu Gln
                2165                2170                2175
Leu Gln Arg Leu Glu Leu Ser Thr Asp Ile Gln Thr Ile Glu Leu
                2180                2185                2190
Gln Ile Lys Lys Leu Lys Glu Leu Gln Lys Ala Val Asp His Arg
```

```
                2195                2200                2205
Lys Ala Ile Ile Leu Ser Ile Asn Leu Cys Ser Pro Glu Phe Thr
                2210                2215                2220
Gln Ala Asp Ser Lys Glu Ser Arg Asp Leu Gln Asp Arg Leu Ser
                2225                2230                2235
Gln Met Asn Gly Arg Trp Asp Arg Val Cys Ser Leu Leu Glu Glu
                2240                2245                2250
Trp Arg Gly Leu Leu Gln Asp Ala Leu Met Gln Cys Gln Gly Phe
                2255                2260                2265
His Glu Met Ser His Gly Leu Leu Leu Met Leu Glu Asn Ile Asp
                2270                2275                2280
Arg Arg Lys Asn Glu Ile Val Pro Ile Asp Ser Asn Leu Asp Ala
                2285                2290                2295
Glu Ile Leu Gln Asp His His Lys Gln Leu Met Gln Ile Lys His
                2300                2305                2310
Glu Leu Leu Glu Ser Gln Leu Arg Val Ala Ser Leu Gln Asp Met
                2315                2320                2325
Ser Cys Gln Leu Leu Val Asn Ala Glu Gly Thr Asp Cys Leu Glu
                2330                2335                2340
Ala Lys Glu Lys Val His Val Ile Gly Asn Arg Leu Lys Leu Leu
                2345                2350                2355
Leu Lys Glu Val Ser Arg His Ile Lys Glu Leu Glu Lys Leu Leu
                2360                2365                2370
Asp Val Ser Ser Ser Gln Gln Asp Leu Ser Ser Trp Ser Ser Ala
                2375                2380                2385
Asp Glu Leu Asp Thr Ser Gly Ser Val Ser Pro Thr Ser Gly Arg
                2390                2395                2400
Ser Thr Pro Asn Arg Gln Lys Thr Pro Arg Gly Lys Cys Ser Leu
                2405                2410                2415
Ser Gln Pro Gly Pro Ser Val Ser Ser Pro His Ser Arg Ser Thr
                2420                2425                2430
Lys Gly Gly Ser Asp Ser Ser Leu Ser Glu Pro Gly Pro Gly Arg
                2435                2440                2445
Ser Gly Arg Gly Phe Leu Phe Arg Val Leu Arg Ala Ala Leu Pro
                2450                2455                2460
Leu Gln Leu Leu Leu Leu Leu Leu Ile Gly Leu Ala Cys Leu Val
                2465                2470                2475
Pro Met Ser Glu Glu Asp Tyr Ser Cys Ala Leu Ser Asn Asn Phe
                2480                2485                2490
Ala Arg Ser Phe His Pro Met Leu Arg Tyr Thr Asn Gly Pro Pro
                2495                2500                2505
Pro Leu
```

<210> 3
<211> 375
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7503554CD1

<400> 3

```
Met Lys Asp Met Pro Leu Arg Ile His Val Leu Leu Gly Leu Ala
  1               5                  10                  15
```

```
Ile Thr Thr Leu Val Gln Ala Val Asp Lys Lys Val Asp Cys Pro
                20                  25                  30
Arg Leu Cys Thr Cys Glu Ile Arg Pro Trp Phe Thr Pro Arg Ser
                35                  40                  45
Ile Tyr Met Glu Ala Ser Thr Val Asp Cys Asn Asp Leu Gly Leu
                50                  55                  60
Leu Thr Phe Pro Ala Arg Leu Pro Ala Asn Thr Gln Ile Leu Leu
                65                  70                  75
Leu Gln Thr Asn Asn Ile Ala Lys Ile Glu Tyr Ser Thr Asp Phe
                80                  85                  90
Pro Val Asn Leu Thr Gly Leu Asp Leu Ser Gln Asn Asn Leu Ser
                95                 100                 105
Ser Val Thr Asn Ile Asn Val Lys Lys Met Pro Gln Leu Leu Ser
               110                 115                 120
Val Tyr Leu Glu Glu Asn Lys Leu Thr Glu Leu Pro Glu Lys Cys
               125                 130                 135
Leu Ser Glu Leu Ser Asn Leu Gln Glu Leu Tyr Ile Asn His Asn
               140                 145                 150
Leu Leu Ser Thr Ile Ser Pro Gly Ala Phe Ile Gly Leu His Asn
               155                 160                 165
Leu Leu Arg Leu His Leu Asn Ser Asn Arg Leu Gln Met Ile Asn
               170                 175                 180
Ser Lys Trp Phe Asp Ala Leu Pro Asn Leu Glu Ile Leu Met Ile
               185                 190                 195
Gly Glu Asn Pro Ile Ile Arg Ile Lys Asp Met Asn Phe Lys Pro
               200                 205                 210
Leu Ile Asn Leu Arg Ser Leu Val Ile Ala Gly Ile Asn Leu Thr
               215                 220                 225
Glu Ile Pro Asp Asn Ala Leu Val Gly Leu Glu Asn Leu Glu Ser
               230                 235                 240
Ile Ser Phe Tyr Asp Asn Arg Leu Ile Lys Val Pro His Val Ala
               245                 250                 255
Leu Gln Lys Val Val Asn Leu Lys Phe Leu Asp Leu Asn Lys Asn
               260                 265                 270
Pro Ile Asn Arg Ile Arg Arg Gly Asp Phe Ser Asn Met Leu His
               275                 280                 285
Leu Lys Glu Leu Gly Ile Asn Asn Met Pro Glu Leu Ile Ser Ile
               290                 295                 300
Asp Ser Leu Ala Val Asp Asn Leu Pro Asp Leu Arg Lys Ile Glu
               305                 310                 315
Ala Thr Asn Asn Pro Arg Leu Ser Tyr Ile His Pro Asn Ala Phe
               320                 325                 330
Phe Arg Leu Pro Lys Leu Glu Ser Leu Met Leu Asn Ser Asn Ala
               335                 340                 345
Leu Ser Ala Leu Cys His Pro Leu Asp Glu His Glu Gln Asn Gln
               350                 355                 360
His Ser Ile His Gly Ala Arg Phe Thr Val Leu Arg Gly Pro Thr
               365                 370                 375
```

<210> 4
<211> 340
<212> PRT
<213> Homo sapiens

<220>

<221> misc_feature
<223> Incyte ID No: 7506077CD1

<400> 4
Met Gly Gly Cys Ala Gly Ser Arg Arg Arg Phe Ser Asp Ser Glu
1 5 10 15
Gly Glu Glu Thr Val Pro Glu Pro Arg Leu Pro Leu Leu Asp His
20 25 30
Gln Gly Ala His Trp Lys Asn Ala Val Gly Phe Trp Leu Leu Gly
35 40 45
Leu Cys Asn Asn Phe Ser Tyr Val Val Met Leu Ser Ala Ala His
50 55 60
Asp Ile Leu Ser His Lys Arg Thr Ser Gly Asn Gln Ser His Val
65 70 75
Asp Pro Gly Pro Thr Pro Ile Pro His Asn Ser Ser Ser Arg Phe
80 85 90
Asp Cys Asn Ser Val Ser Thr Ala Leu Ser Tyr Leu Gly Leu Thr
95 100 105
Gln Ala Gly Leu Ser Pro Gln Gln Thr Leu Leu Ser Met Leu Gly
110 115 120
Ile Pro Ala Leu Leu Leu Ala Ser Tyr Phe Leu Leu Leu Thr Ser
125 130 135
Pro Glu Ala Gln Asp Pro Gly Gly Glu Glu Glu Ala Glu Ser Ala
140 145 150
Ala Arg Gln Pro Leu Ile Arg Thr Glu Ala Pro Glu Ser Lys Pro
155 160 165
Gly Ser Ser Ser Ser Leu Ser Leu Arg Glu Arg Trp Thr Val Phe
170 175 180
Lys Gly Leu Leu Trp Tyr Ile Val Pro Leu Val Val Val Tyr Phe
185 190 195
Ala Glu Tyr Phe Ile Asn Gln Gly Leu Phe Glu Leu Leu Phe Phe
200 205 210
Trp Asn Thr Ser Leu Ser His Ala Gln Gln Tyr Arg Trp Tyr Gln
215 220 225
Met Leu Tyr Gln Ala Gly Val Phe Ala Ser Arg Ser Ser Leu Arg
230 235 240
Cys Cys Arg Ile Arg Phe Thr Trp Ala Leu Ala Leu Leu Gln Cys
245 250 255
Leu Asn Leu Val Phe Leu Leu Ala Asp Val Trp Phe Gly Phe Leu
260 265 270
Pro Ser Ile Tyr Leu Val Phe Leu Ile Ile Leu Tyr Glu Gly Leu
275 280 285
Leu Gly Gly Ala Ala Tyr Val Asn Thr Phe His Asn Ile Ala Leu
290 295 300
Glu Thr Ser Asp Glu His Arg Glu Phe Ala Met Ala Ala Thr Cys
305 310 315
Ile Ser Asp Thr Leu Gly Ile Ser Leu Ser Gly Leu Leu Ala Leu
320 325 330
Pro Leu His Asp Phe Leu Cys Gln Leu Ser
335 340

<210> 5
<211> 287
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7506765CD1

<400> 5
Met Ala Leu Gly Leu Leu Ile Ala Val Pro Leu Leu Leu Gln Ala
1 5 10 15
Ala Pro Arg Gly Ala Ala His Tyr Glu Met Met Gly Thr Cys Arg
20 25 30
Met Ile Cys Asp Pro Tyr Thr Ala Ala Pro Gly Gly Glu Pro Pro
35 40 45
Gly Ala Lys Ala Gln Pro Pro Gly Pro Ser Thr Ala Ala Leu Glu
50 55 60
Val Met Gln Asp Leu Ser Ala Asn Pro Pro Pro Pro Phe Ile Gln
65 70 75
Gly Pro Lys Gly Asp Pro Gly Arg Pro Gly Lys Pro Gly Pro Arg
80 85 90
Gly Pro Pro Gly Glu Pro Gly Pro Pro Gly Pro Arg Gly Pro Pro
95 100 105
Gly Glu Lys Gly Asp Ser Gly Arg Pro Gly Leu Pro Gly Leu Gln
110 115 120
Leu Thr Ala Gly Thr Ala Ser Gly Val Gly Val Val Gly Gly Gly
125 130 135
Ala Gly Val Gly Gly Asp Ser Glu Gly Glu Val Thr Ser Ala Leu
140 145 150
Ser Ala Thr Phe Ser Gly Pro Lys Ile Ala Phe Tyr Val Gly Leu
155 160 165
Lys Ser Pro His Glu Gly Tyr Glu Val Leu Lys Phe Asp Asp Val
170 175 180
Val Thr Asn Leu Gly Asn His Tyr Asp Pro Thr Thr Gly Lys Phe
185 190 195
Ser Cys Gln Val Arg Gly Ile Tyr Phe Phe Thr Tyr His Ile Leu
200 205 210
Met Arg Gly Gly Asp Gly Thr Ser Met Trp Ala Asp Leu Cys Lys
215 220 225
Asn Gly Gln Val Arg Ala Ser Ala Ile Ala Gln Asp Ala Asp Gln
230 235 240
Asn Tyr Asp Tyr Ala Ser Asn Ser Val Val Leu His Leu Asp Ser
245 250 255
Gly Asp Glu Val Tyr Val Lys Leu Asp Gly Gly Lys Ala His Gly
260 265 270
Gly Asn Asn Asn Lys Tyr Ser Thr Phe Ser Gly Phe Leu Leu Tyr
275 280 285
Pro Asp

<210> 6
<211> 1048
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7490048CD1

<400> 6

```
Met Leu Ala Val Ser Leu Lys Trp Arg Leu Gly Val Val Arg Arg
 1               5                  10                  15
Arg Leu Lys Asp Asp Gly Pro Tyr Ser Lys Gly Gly Lys Asp Ala
            20                  25                  30
Gly Gly Ala Asp Val Ser Leu Ala Cys Arg Arg Gln Ser Ile Pro
        35                  40                  45
Glu Glu Phe Arg Gly Ile Thr Val Val Glu Leu Ile Lys Lys Glu
    50                  55                  60
Gly Ser Thr Leu Gly Leu Thr Ile Ser Gly Gly Thr Asp Lys Asp
65                  70                  75
Gly Lys Pro Arg Val Ser Asn Leu Arg Pro Gly Gly Leu Ala Ala
                80                  85                  90
Arg Ser Asp Leu Leu Asn Ile Gly Asp Tyr Ile Arg Ser Val Asn
            95                  100                 105
Gly Ile His Leu Thr Arg Leu Arg His Asp Glu Ile Ile Thr Leu
        110                 115                 120
Leu Lys Asn Val Gly Glu Arg Val Val Leu Glu Val Glu Tyr Glu
    125                 130                 135
Leu Pro Pro Pro Ala Pro Glu Asn Asn Pro Arg Ile Ile Ser Lys
140                 145                 150
Thr Val Asp Val Ser Leu Tyr Lys Glu Gly Asn Ser Phe Gly Phe
155                 160                 165
Val Leu Arg Gly Gly Ala His Glu Asp Gly His Lys Ser Arg Pro
                170                 175                 180
Leu Val Leu Thr Tyr Val Arg Pro Gly Gly Pro Ala Asp Arg Glu
            185                 190                 195
Gly Ser Leu Lys Val Gly Asp Arg Leu Leu Ser Val Asp Gly Ile
        200                 205                 210
Pro Leu His Gly Ala Ser His Ala Thr Ala Leu Ala Thr Leu Arg
    215                 220                 225
Gln Cys Ser His Glu Ala Leu Phe Gln Val Glu Tyr Asp Val Ala
230                 235                 240
Thr Pro Asp Thr Val Ala Asn Ala Ser Gly Pro Leu Met Val Glu
                245                 250                 255
Ile Val Lys Thr Pro Gly Ser Ala Leu Gly Ile Ser Leu Thr Thr
            260                 265                 270
Thr Ser Leu Arg Asn Lys Ser Val Ile Thr Ile Asp Arg Ile Lys
        275                 280                 285
Pro Ala Ser Val Val Asp Arg Ser Gly Ala Leu His Pro Gly Asp
    290                 295                 300
His Ile Leu Ser Ile Asp Gly Thr Ser Met Glu His Cys Ser Leu
305                 310                 315
Leu Glu Ala Thr Lys Leu Leu Ala Ser Ile Ser Glu Lys Val Arg
                320                 325                 330
Leu Glu Ile Leu Pro Val Pro Gln Ser Gln Arg Pro Leu Arg Pro
            335                 340                 345
Ser Glu Ala Val Lys Val Gln Arg Ser Glu Gln Leu His Arg Trp
        350                 355                 360
Asp Pro Cys Val Pro Ser Cys His Ser Pro Arg Pro Gly His Cys
    365                 370                 375
Arg Met Pro Thr Trp Ala Thr Pro Ala Gly Gln Asp Gln Ser Arg
380                 385                 390
Ser Leu Ser Ser Thr Pro Phe Ser Ser Pro Thr Leu Asn His Ala
395                 400                 405
Phe Ser Cys Asn Asn Pro Ser Thr Leu Pro Arg Gly Ser Gln Pro
                410                 415                 420
```

```
Met Ser Pro Arg Thr Thr Met Gly Arg Arg Arg His Arg Arg Arg
                425                 430                 435
Glu His Lys Thr Ser Leu Ser Leu Ala Ser Ser Thr Val Gly Pro
                440                 445                 450
Gly Gly Gln Ile Val His Thr Glu Thr Thr Glu Val Val Leu Cys
                455                 460                 465
Gly Asp Pro Leu Ser Gly Phe Gly Leu Gln Leu Gln Gly Gly Ile
                470                 475                 480
Phe Ala Thr Glu Thr Leu Ser Ser Pro Pro Leu Val Cys Phe Ile
                485                 490                 495
Glu Pro Asp Ser Pro Ala Glu Arg Cys Gly Leu Leu Gln Val Gly
                500                 505                 510
Asp Arg Val Leu Ser Ile Asn Gly Ile Ala Thr Glu Asp Gly Thr
                515                 520                 525
Met Glu Glu Ala Asn Gln Leu Leu Arg Asp Ala Ala Leu Ala His
                530                 535                 540
Lys Val Val Leu Glu Val Glu Phe Asp Val Ala Glu Ser Val Ile
                545                 550                 555
Pro Ser Ser Gly Thr Phe His Val Lys Leu Pro Lys Lys Arg Ser
                560                 565                 570
Val Glu Leu Gly Ile Thr Ile Ser Ser Ala Ser Arg Lys Arg Gly
                575                 580                 585
Glu Pro Leu Ile Ile Ser Asp Ile Lys Lys Gly Ser Val Ala His
                590                 595                 600
Arg Thr Gly Thr Leu Glu Pro Gly Asp Lys Leu Leu Ala Ile Asp
                605                 610                 615
Asn Ile Arg Leu Asp Asn Cys Pro Met Glu Asp Ala Val Gln Ile
                620                 625                 630
Leu Arg Gln Cys Glu Asp Leu Val Lys Leu Lys Ile Arg Lys Asp
                635                 640                 645
Glu Asp Asn Ser Asp Glu Leu Glu Thr Thr Gly Ala Val Ser Tyr
                650                 655                 660
Thr Val Glu Leu Lys Arg Tyr Gly Gly Pro Leu Gly Ile Thr Ile
                665                 670                 675
Ser Gly Thr Glu Glu Pro Phe Asp Pro Ile Val Ile Ser Gly Leu
                680                 685                 690
Thr Lys Arg Gly Leu Ala Glu Arg Thr Gly Ala Ile His Val Gly
                695                 700                 705
Asp Arg Ile Leu Ala Ile Asn Asn Val Ser Leu Lys Gly Arg Pro
                710                 715                 720
Leu Ser Glu Ala Ile His Leu Leu Gln Val Ala Gly Glu Thr Val
                725                 730                 735
Thr Leu Lys Ile Lys Lys Gln Leu Asp Arg Pro Leu Leu Pro Arg
                740                 745                 750
Lys Ser Gly Ser Leu Ser Glu Thr Ser Asp Ala Asp Glu Asp Pro
                755                 760                 765
Ala Asp Ala Leu Lys Gly Gly Leu Pro Ala Ala Arg Phe Ser Pro
                770                 775                 780
Ala Val Pro Ser Val Asp Ser Ala Val Glu Ser Trp Asp Ser Ser
                785                 790                 795
Ala Thr Glu Gly Gly Phe Gly Gly Pro Gly Ser Tyr Thr Pro Gln
                800                 805                 810
Ala Ala Ala Arg Gly Thr Thr Pro Gln Glu Arg Arg Pro Gly Trp
                815                 820                 825
Leu Arg Gly Ser Pro Pro Pro Thr Glu Pro Arg Arg Thr Ser Tyr
                830                 835                 840
```

```
Thr Pro Thr Pro Ala Asp Glu Ser Phe Pro Glu Glu Glu Glu Glu
                845                 850                 855
Asp Asp Trp Glu Pro Pro Thr Ser Pro Ala Pro Gly Pro Ala Arg
                860                 865                 870
Glu Glu Gly Phe Trp Arg Met Phe Gly Glu Ala Leu Glu Asp Leu
                875                 880                 885
Glu Ser Cys Gly Gln Ser Glu Leu Leu Arg Glu Leu Glu Ala Ser
                890                 895                 900
Ile Met Thr Gly Thr Val Gln Arg Val Ala Leu Glu Gly Arg Pro
                905                 910                 915
Gly His Arg Pro Trp Gln Arg Gly Arg Glu Val Arg Ala Ser Pro
                920                 925                 930
Ala Glu Met Glu Glu Leu Leu Leu Pro Thr Pro Leu Glu Met His
                935                 940                 945
Lys Val Thr Leu His Lys Asp Pro Met Arg His Asp Phe Gly Phe
                950                 955                 960
Ser Val Ser Asp Gly Leu Leu Glu Lys Gly Val Tyr Val His Thr
                965                 970                 975
Val Arg Pro Asp Gly Pro Ala His Arg Gly Gly Leu Gln Pro Phe
                980                 985                 990
Asp Arg Val Leu Gln Val Asn His Val Arg Thr Arg Asp Phe Asp
                995                1000                1005
Cys Cys Leu Ala Val Pro Leu Leu Ala Glu Ala Gly Asp Val Leu
               1010                1015                1020
Glu Leu Ile Ile Ser Arg Lys Pro His Thr Ala His Ser Ser Arg
               1025                1030                1035
Ala Pro Arg Ser Pro Gly Pro Ser Ser Pro Arg Met Leu
               1040                1045
```

<210> 7
<211> 419
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 2133585CD1

<400> 7

```
Met Gly Ser Pro Ala His Arg Pro Ala Leu Leu Leu Leu Leu Pro
  1               5                  10                  15
Pro Leu Leu Leu Leu Leu Leu Leu Arg Val Pro Pro Ser Arg Ser
                20                  25                  30
Phe Pro Gly Ser Gly Asp Ser Pro Leu Glu Asp Asp Glu Val Gly
                35                  40                  45
Tyr Ser His Pro Arg Tyr Lys Asp Thr Pro Trp Cys Ser Pro Ile
                50                  55                  60
Lys Val Lys Tyr Gly Asp Val Tyr Cys Arg Ala Pro Gln Gly Gly
                65                  70                  75
Tyr Tyr Lys Thr Ala Leu Gly Thr Arg Cys Asp Ile Arg Cys Gln
                80                  85                  90
Lys Gly Tyr Glu Leu His Gly Ser Ser Leu Leu Ile Cys Gln Ser
                95                 100                 105
Asn Lys Arg Trp Ser Asp Lys Val Ile Cys Arg Gln Lys Arg Cys
               110                 115                 120
Pro Thr Leu Ala Met Pro Ala Asn Gly Gly Phe Lys Cys Val Asp
```

```
                125                 130                 135
Gly Ala Tyr Phe Asn Ser Arg Cys Glu Tyr Tyr Cys Ser Pro Gly
                140                 145                 150
Tyr Thr Leu Lys Gly Glu Arg Thr Val Thr Cys Met Asp Asn Lys
                155                 160                 165
Ala Trp Ser Gly Arg Pro Ala Ser Cys Val Asp Met Glu Pro Pro
                170                 175                 180
Arg Ile Lys Cys Pro Ser Val Lys Glu Arg Ile Ala Glu Pro Asn
                185                 190                 195
Lys Leu Thr Val Arg Val Ser Trp Glu Thr Pro Glu Gly Arg Asp
                200                 205                 210
Thr Ala Asp Gly Ile Leu Thr Asp Val Ile Leu Lys Gly Leu Pro
                215                 220                 225
Pro Gly Ser Asn Phe Pro Glu Gly Asp His Lys Ile Gln Tyr Thr
                230                 235                 240
Val Tyr Asp Arg Ala Glu Asn Lys Gly Thr Cys Lys Phe Arg Val
                245                 250                 255
Lys Val Arg Val Lys Arg Cys Gly Lys Leu Asn Ala Pro Glu Asn
                260                 265                 270
Gly Tyr Met Lys Cys Ser Ser Asp Gly Asp Asn Tyr Gly Ala Thr
                275                 280                 285
Cys Glu Phe Ser Cys Ile Gly Gly Tyr Glu Leu Gln Gly Ser Pro
                290                 295                 300
Ala Arg Val Cys Gln Ser Asn Leu Ala Trp Ser Gly Thr Glu Pro
                305                 310                 315
Thr Cys Ala Ala Met Asn Val Asn Val Gly Val Arg Thr Ala Ala
                320                 325                 330
Ala Leu Leu Asp Gln Phe Tyr Glu Lys Arg Arg Leu Leu Ile Val
                335                 340                 345
Ser Thr Pro Thr Ala Arg Asn Leu Leu Tyr Arg Leu Gln Leu Gly
                350                 355                 360
Met Leu Gln Gln Ala Gln Cys Gly Leu Asp Leu Arg His Ile Thr
                365                 370                 375
Val Val Glu Leu Val Gly Val Phe Pro Thr Leu Ile Gly Arg Ile
                380                 385                 390
Gly Ala Lys Ile Met Pro Pro Ala Leu Ala Leu Gln Leu Ser Phe
                395                 400                 405
Ile Thr Leu Thr Lys Leu Glu Pro Trp Ile Leu Val Ser Glu
                410                 415
```

<210> 8
<211> 1109
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7509063CD1

<400> 8

```
Met Arg Cys Lys Glu Leu Glu Asn Ala Val Gly Ser Trp Thr Asp
  1               5                  10                  15
Asp Leu Thr Gln Leu Ser Leu Leu Lys Asp Thr Leu Ser Ala Tyr
                20                  25                  30
Ile Ser Ala Asp Asp Ile Ser Ile Leu Asn Glu Arg Val Glu Leu
                35                  40                  45
```

```
Leu Gln Arg Gln Trp Glu Glu Leu Cys His Gln Leu Ser Leu Arg
                50                  55                  60
Arg Gln Gln Ile Gly Glu Arg Leu Asn Glu Trp Ala Val Phe Ser
                65                  70                  75
Glu Lys Asn Lys Glu Leu Cys Glu Trp Leu Thr Gln Met Glu Ser
                80                  85                  90
Lys Val Ser Gln Asn Gly Asp Ile Leu Ile Glu Glu Met Ile Glu
                95                 100                 105
Lys Leu Lys Lys Asp Tyr Gln Glu Glu Ile Ala Ile Ala Gln Glu
               110                 115                 120
Asn Lys Ile Gln Leu Gln Gln Met Gly Glu Arg Leu Ala Lys Ala
               125                 130                 135
Ser His Glu Ser Lys Ala Ser Glu Ile Glu Tyr Lys Leu Gly Lys
               140                 145                 150
Val Asn Asp Arg Trp Gln His Leu Leu Asp Leu Ile Ala Ala Arg
               155                 160                 165
Val Lys Lys Leu Lys Glu Thr Leu Val Ala Val Gln Gln Leu Asp
               170                 175                 180
Lys Asn Met Ser Ser Leu Arg Thr Trp Leu Ala His Ile Glu Ser
               185                 190                 195
Glu Leu Ala Lys Pro Ile Val Tyr Asp Ser Cys Asn Ser Glu Glu
               200                 205                 210
Ile Gln Arg Lys Leu Asn Glu Gln Gln Glu Leu Gln Arg Asp Ile
               215                 220                 225
Glu Lys His Ser Thr Gly Val Ala Ser Val Leu Asn Leu Cys Glu
               230                 235                 240
Val Leu Leu His Asp Cys Asp Ala Cys Ala Thr Asp Ala Glu Cys
               245                 250                 255
Asp Ser Ile Gln Gln Ala Thr Arg Asn Leu Asp Arg Arg Trp Arg
               260                 265                 270
Asn Ile Cys Ala Met Ser Met Glu Arg Arg Leu Lys Ile Glu Glu
               275                 280                 285
Thr Trp Arg Leu Trp Gln Lys Phe Leu Asp Asp Tyr Ser Arg Phe
               290                 295                 300
Glu Asp Trp Leu Lys Ser Ser Glu Arg Thr Ala Ala Phe Pro Ser
               305                 310                 315
Ser Ser Gly Val Ile Tyr Thr Val Ala Lys Glu Glu Leu Lys Lys
               320                 325                 330
Phe Glu Ala Phe Gln Arg Gln Val His Glu Cys Leu Thr Gln Leu
               335                 340                 345
Glu Leu Ile Asn Lys Gln Tyr Arg Arg Leu Ala Arg Glu Asn Arg
               350                 355                 360
Thr Asp Ser Ala Cys Ser Leu Lys Gln Met Val His Glu Gly Asn
               365                 370                 375
Gln Arg Trp Asp Asn Leu Gln Lys Arg Val Thr Ser Ile Leu Arg
               380                 385                 390
Arg Leu Lys His Phe Ile Gly Gln Arg Glu Glu Phe Glu Thr Ala
               395                 400                 405
Arg Asp Ser Ile Leu Val Trp Leu Thr Glu Met Asp Leu Gln Leu
               410                 415                 420
Thr Asn Ile Glu His Phe Ser Glu Cys Asp Val Gln Ala Lys Ile
               425                 430                 435
Lys Gln Leu Lys Ala Phe Gln Gln Glu Ile Ser Leu Asn His Asn
               440                 445                 450
Lys Ile Glu Gln Ile Ile Ala Gln Gly Glu Gln Leu Ile Glu Lys
               455                 460                 465
```

```
Ser Glu Pro Leu Asp Ala Ala Ile Ile Glu Glu Glu Leu Asp Glu
                470                 475                 480
Leu Arg Arg Tyr Cys Gln Glu Val Phe Gly Arg Val Glu Arg Tyr
                485                 490                 495
His Lys Lys Leu Ile Arg Leu Pro Leu Pro Asp Asp Glu His Asp
                500                 505                 510
Leu Ser Asp Arg Glu Leu Glu Leu Glu Asp Ser Ala Ala Leu Ser
                515                 520                 525
Asp Leu His Trp His Asp Arg Ser Ala Asp Ser Leu Leu Ser Pro
                530                 535                 540
Gln Pro Ser Ser Asn Leu Ser Leu Ser Leu Ala Gln Pro Leu Arg
                545                 550                 555
Ser Glu Arg Ser Gly Arg Asp Thr Pro Ala Ser Val Asp Ser Ile
                560                 565                 570
Pro Leu Glu Trp Asp His Asp Tyr Asp Leu Ser Arg Asp Leu Glu
                575                 580                 585
Ser Ala Met Ser Arg Ala Leu Pro Ser Glu Asp Glu Glu Gly Gln
                590                 595                 600
Asp Asp Lys Asp Phe Tyr Leu Arg Gly Ala Val Ala Leu Ser Asp
                605                 610                 615
Val Met Ile Pro Glu Ser Pro Glu Ala Tyr Val Lys Leu Thr Glu
                620                 625                 630
Asn Ala Ile Lys Asn Thr Ser Gly Asp His Ser Ala Leu Glu Ser
                635                 640                 645
Gln Ile Arg Gln Leu Gly Lys Ala Leu Asp Asp Ser Arg Phe Gln
                650                 655                 660
Ile Gln Gln Thr Glu Asn Ile Ile Arg Ser Lys Thr Pro Thr Gly
                665                 670                 675
Pro Glu Leu Asp Thr Ser Tyr Lys Gly Tyr Met Lys Leu Leu Gly
                680                 685                 690
Glu Cys Ser Ser Ser Ile Asp Ser Val Lys Arg Leu Glu His Lys
                695                 700                 705
Leu Lys Glu Glu Glu Glu Ser Leu Pro Gly Phe Val Asn Leu His
                710                 715                 720
Ser Thr Glu Thr Gln Thr Ala Gly Val Ile Asp Arg Trp Glu Leu
                725                 730                 735
Leu Gln Ala Gln Ala Leu Ser Lys Glu Leu Arg Met Lys Gln Asn
                740                 745                 750
Leu Gln Lys Trp Gln Gln Phe Asn Ser Asp Leu Asn Ser Ile Trp
                755                 760                 765
Ala Trp Leu Gly Asp Thr Glu Glu Glu Leu Glu Gln Leu Gln Arg
                770                 775                 780
Leu Glu Leu Ser Thr Asp Ile Gln Thr Ile Glu Leu Gln Ile Lys
                785                 790                 795
Lys Leu Lys Glu Leu Gln Lys Ala Val Asp His Arg Lys Ala Ile
                800                 805                 810
Ile Leu Ser Ile Asn Leu Cys Ser Pro Glu Phe Thr Gln Ala Asp
                815                 820                 825
Ser Lys Glu Ser Arg Asp Leu Gln Asp Arg Leu Ser Gln Met Asn
                830                 835                 840
Gly Arg Trp Asp Arg Val Cys Ser Leu Leu Glu Glu Trp Arg Gly
                845                 850                 855
Leu Leu Gln Asp Ala Leu Met Gln Cys Gln Gly Phe His Glu Met
                860                 865                 870
Ser His Gly Leu Leu Leu Met Leu Glu Asn Ile Asp Arg Arg Lys
                875                 880                 885
```

```
Asn Glu Ile Val Pro Ile Asp Ser Asn Leu Asp Ala Glu Ile Leu
                890                 895                 900
Gln Asp His His Lys Gln Leu Met Gln Ile Lys His Glu Leu Leu
                905                 910                 915
Glu Ser Gln Leu Arg Val Ala Ser Leu Gln Asp Met Ser Cys Gln
                920                 925                 930
Leu Leu Val Asn Ala Glu Gly Thr Asp Cys Leu Glu Ala Lys Glu
                935                 940                 945
Lys Val His Val Ile Gly Asn Arg Leu Lys Leu Leu Leu Lys Glu
                950                 955                 960
Val Ser Arg His Ile Lys Glu Leu Glu Lys Leu Leu Asp Val Ser
                965                 970                 975
Ser Ser Gln Gln Asp Leu Ser Ser Trp Ser Ser Ala Asp Glu Leu
                980                 985                 990
Asp Thr Ser Gly Ser Val Ser Pro Thr Ser Gly Arg Ser Thr Pro
                995                1000                1005
Asn Arg Gln Lys Thr Pro Arg Gly Lys Cys Ser Leu Ser Gln Pro
               1010                1015                1020
Gly Pro Ser Val Ser Ser Pro His Ser Arg Ser Thr Lys Gly Gly
               1025                1030                1035
Ser Asp Ser Ser Leu Ser Glu Pro Gly Pro Gly Arg Ser Gly Arg
               1040                1045                1050
Gly Phe Leu Phe Arg Val Leu Arg Ala Ala Leu Pro Leu Gln Leu
               1055                1060                1065
Leu Leu Leu Leu Leu Ile Gly Leu Ala Cys Leu Val Pro Met Ser
               1070                1075                1080
Glu Glu Asp Tyr Ser Cys Ala Leu Ser Asn Asn Phe Ala Arg Ser
               1085                1090                1095
Phe His Pro Met Leu Arg Tyr Thr Asn Gly Pro Pro Pro Leu
               1100                1105
```

<210> 9
<211> 1308
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 5496710CD1

<400> 9

```
Met Gly Ser Val Thr Gly Ala Val Leu Lys Thr Leu Leu Leu Leu
  1               5                  10                  15
Ser Thr Gln Asn Trp Asn Arg Val Glu Ala Gly Asn Ser Tyr Asp
                20                  25                  30
Cys Asp Asp Pro Leu Val Ser Ala Leu Pro Gln Ala Ser Phe Ser
                35                  40                  45
Ser Ser Ser Glu Leu Ser Ser Ser His Gly Pro Gly Phe Ala Arg
                50                  55                  60
Leu Asn Arg Arg Asp Gly Ala Gly Gly Trp Ser Pro Leu Val Ser
                65                  70                  75
Asn Lys Tyr Gln Trp Leu Gln Ile Asp Leu Gly Glu Arg Met Glu
                80                  85                  90
Val Thr Ala Val Ala Thr Gln Gly Gly Tyr Gly Ser Ser Asn Trp
                95                 100                 105
Val Thr Ser Tyr Leu Leu Met Phe Ser Asp Ser Gly Trp Asn Trp
```

```
                110                 115                 120
Lys Gln Tyr Arg Gln Glu Asp Ser Ile Trp Gly Phe Ser Gly Asn
                125                 130                 135
Ala Asn Ala Asp Ser Val Val Tyr Tyr Arg Leu Gln Pro Ser Ile
                140                 145                 150
Lys Ala Arg Phe Leu Arg Phe Ile Pro Leu Glu Trp Asn Pro Lys
                155                 160                 165
Gly Arg Ile Gly Met Arg Ile Glu Val Phe Gly Cys Ala Tyr Arg
                170                 175                 180
Ser Glu Val Val Asp Leu Asp Gly Lys Ser Ser Leu Leu Tyr Arg
                185                 190                 195
Phe Asp Gln Lys Ser Leu Ser Pro Ile Lys Asp Ile Ile Ser Leu
                200                 205                 210
Lys Phe Lys Thr Met Gln Ser Asp Gly Ile Leu Leu His Arg Glu
                215                 220                 225
Gly Pro Asn Gly Asp His Ile Thr Leu Gln Leu Arg Arg Ala Arg
                230                 235                 240
Leu Phe Leu Leu Ile Asn Ser Gly Glu Ala Lys Leu Pro Ser Thr
                245                 250                 255
Ser Thr Leu Val Asn Leu Thr Leu Gly Ser Leu Leu Asp Asp Gln
                260                 265                 270
His Trp His Ser Val Leu Ile Gln Arg Leu Gly Lys Gln Val Asn
                275                 280                 285
Phe Thr Val Asp Glu His Arg His His Phe His Ala Arg Gly Glu
                290                 295                 300
Phe Asn Leu Met Asn Leu Asp Tyr Glu Ile Ser Phe Gly Gly Ile
                305                 310                 315
Pro Ala Pro Gly Lys Ser Val Ser Phe Pro His Arg Asn Phe His
                320                 325                 330
Gly Cys Leu Glu Asn Leu Tyr Tyr Asn Gly Val Asp Ile Ile Asp
                335                 340                 345
Leu Ala Lys Gln Gln Lys Pro Gln Ile Ile Ala Met Gly Asn Val
                350                 355                 360
Ser Phe Ser Cys Ser Gln Pro Gln Ser Met Pro Val Thr Phe Leu
                365                 370                 375
Ser Ser Arg Ser Tyr Leu Ala Leu Pro Asp Phe Ser Gly Glu Glu
                380                 385                 390
Glu Val Ser Ala Thr Phe Gln Phe Arg Thr Trp Asn Lys Ala Gly
                395                 400                 405
Leu Leu Leu Phe Ser Glu Leu Gln Leu Ile Ser Gly Gly Ile Leu
                410                 415                 420
Leu Phe Leu Ser Asp Gly Lys Leu Lys Ser Asn Leu Tyr Gln Pro
                425                 430                 435
Gly Lys Leu Pro Ser Asp Ile Thr Ala Gly Val Glu Leu Asn Asp
                440                 445                 450
Gly Gln Trp His Ser Val Ser Leu Ser Ala Lys Lys Asn His Leu
                455                 460                 465
Ser Val Ala Val Asp Gly Gln Met Ala Ser Ala Ala Pro Leu Leu
                470                 475                 480
Gly Pro Glu Gln Ile Tyr Ser Gly Gly Thr Tyr Tyr Phe Gly Gly
                485                 490                 495
Cys Pro Asp Lys Ser Phe Gly Ser Lys Cys Lys Ser Pro Leu Gly
                500                 505                 510
Gly Phe Gln Gly Cys Met Arg Leu Ile Ser Ile Ser Gly Lys Val
                515                 520                 525
Val Asp Leu Ile Ser Val Gln Gln Gly Ser Leu Gly Asn Phe Ser
```

```
                530                 535                 540
Asp Leu Gln Ile Asp Ser Cys Gly Ile Ser Asp Arg Cys Leu Pro
                545                 550                 555
Asn Tyr Cys Glu His Gly Gly Glu Cys Ser Gln Ser Trp Ser Thr
                560                 565                 570
Phe His Cys Asn Cys Thr Asn Thr Gly Tyr Arg Gly Ala Thr Cys
                575                 580                 585
His Asn Ser Ile Tyr Glu Gln Ser Cys Glu Ala Tyr Lys His Arg
                590                 595                 600
Gly Asn Thr Ser Gly Phe Tyr Tyr Ile Asp Ser Asp Gly Ser Gly
                605                 610                 615
Pro Leu Glu Pro Phe Leu Leu Tyr Cys Asn Met Thr Glu Thr Ala
                620                 625                 630
Trp Thr Ile Ile Gln His Asn Gly Ser Asp Leu Thr Arg Val Arg
                635                 640                 645
Asn Thr Asn Pro Glu Asn Pro Tyr Ala Gly Phe Phe Glu Tyr Val
                650                 655                 660
Ala Ser Met Glu Gln Leu Gln Ala Thr Ile Asn Arg Ala Glu His
                665                 670                 675
Cys Glu Gln Glu Phe Thr Tyr Tyr Cys Lys Lys Ser Arg Leu Val
                680                 685                 690
Asn Lys Gln Asp Gly Thr Pro Leu Ser Trp Trp Val Gly Arg Thr
                695                 700                 705
Asn Glu Thr Gln Thr Tyr Trp Gly Gly Ser Ser Pro Asp Leu Gln
                710                 715                 720
Lys Cys Thr Cys Gly Leu Glu Gly Asn Cys Ile Asp Ser Gln Tyr
                725                 730                 735
Tyr Cys Asn Cys Asp Ala Asp Arg Asn Glu Trp Thr Asn Asp Thr
                740                 745                 750
Gly Leu Leu Ala Tyr Lys Glu His Leu Pro Val Thr Lys Ile Val
                755                 760                 765
Ile Thr Asp Thr Gly Arg Leu His Ser Glu Ala Ala Tyr Lys Leu
                770                 775                 780
Gly Pro Leu Leu Cys Arg Gly Asp Arg Ser Phe Trp Asn Ser Ala
                785                 790                 795
Ser Phe Asp Thr Glu Ala Ser Tyr Leu His Phe Pro Thr Phe His
                800                 805                 810
Gly Glu Leu Ser Ala Asp Val Ser Phe Phe Phe Lys Thr Thr Ala
                815                 820                 825
Ser Ser Gly Val Phe Leu Glu Asn Leu Gly Ile Ala Asp Phe Ile
                830                 835                 840
Arg Ile Glu Leu Arg Ser Pro Thr Val Val Thr Phe Ser Phe Asp
                845                 850                 855
Val Gly Asn Gly Pro Phe Glu Ile Ser Val Gln Ser Pro Thr His
                860                 865                 870
Phe Asn Asp Asn Gln Trp His His Val Arg Val Glu Arg Asn Met
                875                 880                 885
Lys Glu Ala Ser Leu Gln Val Asp Gln Leu Thr Pro Lys Thr Gln
                890                 895                 900
Pro Ala Pro Ala Asp Gly His Val Leu Leu Gln Leu Asn Ser Gln
                905                 910                 915
Leu Phe Val Gly Gly Thr Ala Thr Arg Gln Arg Gly Phe Leu Gly
                920                 925                 930
Cys Ile Arg Ser Leu Gln Leu Asn Gly Met Thr Leu Asp Leu Glu
                935                 940                 945
Glu Arg Ala Gln Val Thr Pro Glu Val Gln Pro Gly Cys Arg Gly
```

```
                950                 955                 960
His Cys Ser Ser Tyr Gly Lys Leu Cys Arg Asn Gly Gly Lys Cys
                965                 970                 975
Arg Glu Arg Pro Ile Gly Phe Phe Cys Asp Cys Thr Phe Ser Ala
                980                 985                 990
Tyr Thr Gly Pro Phe Cys Ser Asn Glu Ile Ser Ala Tyr Phe Gly
                995                1000                1005
Ser Gly Ser Ser Val Ile Tyr Asn Phe Gln Glu Asn Tyr Leu Leu
               1010                1015                1020
Ser Lys Asn Ser Ser Ser His Ala Ala Ser Phe His Gly Asp Met
               1025                1030                1035
Lys Leu Ser Arg Glu Met Ile Lys Phe Ser Phe Arg Thr Thr Arg
               1040                1045                1050
Thr Pro Ser Leu Leu Leu Phe Val Ser Ser Phe Tyr Lys Glu Tyr
               1055                1060                1065
Leu Ser Val Ile Ile Ala Lys Asn Gly Ser Leu Gln Ile Arg Tyr
               1070                1075                1080
Lys Leu Asn Lys Tyr Gln Glu Pro Asp Val Val Asn Phe Asp Phe
               1085                1090                1095
Lys Asn Met Ala Asp Gly Gln Leu His His Ile Met Ile Asn Arg
               1100                1105                1110
Glu Glu Gly Val Val Phe Ile Glu Ile Asp Asp Asn Arg Arg Arg
               1115                1120                1125
Gln Val His Leu Ser Ser Gly Thr Glu Phe Ser Ala Val Lys Ser
               1130                1135                1140
Leu Val Leu Gly Arg Ile Leu Glu His Ser Asp Val Asp Gln Asp
               1145                1150                1155
Thr Ala Leu Ala Gly Ala Gln Gly Phe Thr Gly Cys Leu Ser Ala
               1160                1165                1170
Val Gln Leu Ser His Val Ala Pro Leu Lys Ala Ala Leu His Pro
               1175                1180                1185
Ser His Pro Asp Pro Val Thr Val Thr Gly His Val Thr Glu Ser
               1190                1195                1200
Ser Cys Met Ala Gln Pro Gly Thr Asp Ala Thr Ser Arg Glu Arg
               1205                1210                1215
Thr His Ser Phe Ala Asp His Ser Gly Thr Ile Asp Asp Arg Glu
               1220                1225                1230
Pro Leu Ala Asn Ala Ile Lys Ser Asp Ser Ala Val Ile Gly Gly
               1235                1240                1245
Leu Ile Ala Val Val Ile Phe Ile Leu Leu Cys Ile Thr Ala Ile
               1250                1255                1260
Ala Val Arg Ile Tyr Gln Gln Lys Arg Leu Tyr Lys Arg Ser Glu
               1265                1270                1275
Ala Lys Arg Ser Glu Asn Val Asp Ser Ala Glu Ala Val Leu Lys
               1280                1285                1290
Ser Glu Leu Asn Ile Gln Asn Ala Val Asn Glu Asn Gln Lys Glu
               1295                1300                1305
Tyr Phe Phe
```

<210> 10
<211> 1009
<212> PRT
<213> Homo sapiens

<220>

<221> misc_feature
<223> Incyte ID No: 72150826CD1

<400> 10

```
Met Asp Leu Ile Arg Gly Val Leu Leu Arg Leu Leu Leu Leu Ala
  1               5                  10                  15
Ser Ser Leu Gly Pro Gly Ala Val Ser Leu Arg Ala Ala Ile Arg
                20                  25                  30
Lys Pro Gly Lys Val Gly Pro Pro Leu Asp Ile Lys Leu Gly Ala
                35                  40                  45
Leu Asn Cys Thr Ala Phe Ser Ile Gln Trp Lys Met Pro Arg His
                50                  55                  60
Pro Gly Ser Pro Ile Leu Gly Tyr Thr Val Phe Tyr Ser Glu Val
                65                  70                  75
Gly Ala Asp Lys Ser Leu Gln Glu Gln Leu His Ser Val Pro Leu
                80                  85                  90
Ser Arg Asp Ile Pro Thr Thr Glu Glu Val Ile Gly Asp Leu Lys
                95                 100                 105
Pro Gly Thr Glu Tyr His Val Ser Ile Ala Ala Tyr Ser Gln Ala
               110                 115                 120
Gly Lys Gly Arg Leu Ser Ser Pro Arg His Val Thr Thr Leu Ser
               125                 130                 135
Gln Asp Ser Cys Leu Pro Pro Ala Ala Pro Gln Gln Pro His Val
               140                 145                 150
Ile Val Val Ser Asp Ser Glu Val Ala Leu Ser Trp Lys Pro Gly
               155                 160                 165
Ala Ser Glu Gly Ser Ala Pro Ile Gln Tyr Tyr Ser Val Glu Phe
               170                 175                 180
Ile Arg Pro Asp Phe Asp Lys Lys Trp Thr Ser Ile His Glu Arg
               185                 190                 195
Ile Gln Met Asp Ser Met Val Ile Lys Gly Leu Asp Pro Asp Thr
               200                 205                 210
Asn Tyr Gln Phe Ala Val Arg Ala Met Asn Ser His Gly Pro Ser
               215                 220                 225
Pro Arg Ser Trp Pro Ser Asp Ile Ile Arg Thr Leu Cys Pro Glu
               230                 235                 240
Glu Ala Gly Ser Gly Arg Tyr Gly Pro Arg Tyr Ile Thr Asp Met
               245                 250                 255
Gly Ala Gly Glu Asp Asp Glu Gly Phe Glu Asp Asp Leu Asp Leu
               260                 265                 270
Asp Ile Ser Phe Glu Glu Val Lys Pro Leu Pro Ala Thr Lys Gly
               275                 280                 285
Gly Asn Lys Lys Phe Leu Val Glu Ser Lys Lys Met Ser Ile Ser
               290                 295                 300
Asn Pro Lys Thr Ile Ser Arg Leu Ile Pro Pro Thr Ser Ala Ser
               305                 310                 315
Leu Pro Val Thr Thr Val Ala Pro Gln Pro Ile Pro Ile Gln Arg
               320                 325                 330
Lys Gly Lys Asn Gly Val Ala Ile Met Ser Arg Leu Phe Asp Met
               335                 340                 345
Pro Cys Asp Glu Thr Leu Cys Ser Ala Asp Ser Phe Cys Val Asn
               350                 355                 360
Asp Tyr Thr Trp Gly Gly Ser Arg Cys Gln Cys Thr Leu Gly Lys
               365                 370                 375
Gly Gly Glu Ser Cys Ser Glu Asp Ile Val Ile Gln Tyr Pro Gln
               380                 385                 390
```

```
Phe Phe Gly His Ser Tyr Val Thr Phe Glu Pro Leu Lys Asn Ser
                395                 400                 405
Tyr Gln Ala Phe Gln Ile Thr Leu Glu Phe Arg Ala Glu Ala Glu
                410                 415                 420
Asp Gly Leu Leu Leu Tyr Cys Gly Glu Asn Glu His Gly Arg Gly
                425                 430                 435
Asp Phe Met Ser Leu Ala Ile Ile Arg Arg Ser Leu Gln Phe Arg
                440                 445                 450
Phe Asn Cys Gly Thr Gly Val Ala Ile Ile Val Ser Glu Thr Lys
                455                 460                 465
Ile Lys Leu Gly Gly Trp His Thr Val Met Leu Tyr Arg Asp Gly
                470                 475                 480
Leu Asn Gly Leu Leu Gln Leu Asn Asn Gly Thr Pro Val Thr Gly
                485                 490                 495
Gln Ser Gln Gly Gln Tyr Ser Lys Ile Thr Phe Arg Thr Pro Leu
                500                 505                 510
Tyr Leu Gly Gly Ala Pro Ser Ala Tyr Trp Leu Val Arg Ala Thr
                515                 520                 525
Gly Thr Asn Arg Gly Phe Gln Gly Cys Val Gln Ser Leu Ala Val
                530                 535                 540
Asn Gly Arg Arg Ile Asp Met Arg Pro Trp Pro Leu Gly Lys Ala
                545                 550                 555
Leu Ser Gly Ala Asp Val Gly Glu Cys Ser Ser Gly Ile Cys Asp
                560                 565                 570
Glu Ala Ser Cys Ile His Gly Gly Thr Cys Thr Ala Ile Lys Ala
                575                 580                 585
Asp Ser Tyr Ile Cys Leu Cys Pro Leu Gly Phe Lys Gly Arg His
                590                 595                 600
Cys Glu Asp Ala Phe Thr Leu Thr Ile Pro Gln Phe Arg Glu Ser
                605                 610                 615
Leu Arg Ser Tyr Ala Ala Thr Pro Trp Pro Leu Glu Pro Gln His
                620                 625                 630
Tyr Leu Ser Phe Met Glu Phe Glu Ile Thr Phe Arg Pro Asp Ser
                635                 640                 645
Gly Asp Gly Val Leu Leu Tyr Ser Tyr Asp Thr Gly Ser Lys Asp
                650                 655                 660
Phe Leu Ser Ile Asn Leu Ala Gly Gly His Val Glu Phe Arg Phe
                665                 670                 675
Asp Cys Gly Ser Gly Thr Gly Val Leu Arg Ser Glu Asp Pro Leu
                680                 685                 690
Thr Leu Gly Asn Trp His Glu Leu Arg Val Ser Arg Thr Ala Lys
                695                 700                 705
Asn Gly Ile Leu Gln Val Asp Lys Gln Lys Ile Val Glu Gly Met
                710                 715                 720
Ala Glu Gly Gly Phe Thr Gln Ile Lys Cys Asn Thr Asp Ile Phe
                725                 730                 735
Ile Gly Gly Val Pro Asn Tyr Asp Asp Val Lys Lys Asn Ser Gly
                740                 745                 750
Val Leu Lys Pro Phe Ser Gly Ser Ile Gln Lys Ile Ile Leu Asn
                755                 760                 765
Asp Arg Thr Ile His Val Lys His Asp Phe Thr Ser Gly Val Asn
                770                 775                 780
Val Glu Asn Ala Ala His Pro Cys Val Arg Ala Pro Cys Ala His
                785                 790                 795
Gly Gly Ser Cys Arg Pro Arg Lys Glu Gly Tyr Asp Cys Asp Cys
                800                 805                 810
```

```
Pro Leu Gly Phe Glu Gly Leu His Cys Gln Lys Ala Ile Ile Glu
                815                 820                 825
Ala Ile Glu Ile Pro Gln Phe Ile Gly Arg Ser Tyr Leu Thr Tyr
                830                 835                 840
Asp Asn Pro Asp Ile Leu Lys Arg Val Ser Gly Ser Arg Ser Asn
                845                 850                 855
Val Phe Met Arg Phe Lys Thr Thr Ala Lys Asp Gly Leu Leu Leu
                860                 865                 870
Trp Arg Gly Asp Ser Pro Met Arg Pro Asn Ser Asp Phe Ile Ser
                875                 880                 885
Leu Gly Leu Arg Asp Gly Ala Leu Val Phe Ser Tyr Asn Leu Gly
                890                 895                 900
Ser Gly Val Ala Ser Ile Met Val Asn Gly Ser Phe Asn Asp Gly
                905                 910                 915
Arg Trp His Arg Val Lys Ala Val Arg Asp Gly Gln Ser Gly Lys
                920                 925                 930
Ile Thr Val Asp Asp Tyr Gly Ala Arg Thr Gly Lys Ser Pro Gly
                935                 940                 945
Met Met Arg Gln Leu Asn Ile Asn Gly Ala Leu Tyr Val Gly Gly
                950                 955                 960
Met Lys Glu Ile Ala Leu His Thr Asn Arg Gln Tyr Met Arg Gly
                965                 970                 975
Leu Val Gly Cys Ile Ser His Phe Thr Leu Ser Thr Asp Tyr His
                980                 985                 990
Ile Ser Leu Val Glu Asp Ala Val Asp Gly Lys Asn Ile Asn Thr
                995                1000                1005
Cys Gly Ala Lys
```

<210> 11
<211> 962
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 6799476CD1

<400> 11

```
Met Leu Arg Arg Pro Ala Pro Ala Leu Ala Pro Ala Ala Arg Leu
  1               5                  10                  15
Leu Leu Ala Gly Leu Leu Cys Gly Gly Gly Val Trp Ala Ala Arg
                20                  25                  30
Val Asn Lys His Lys Pro Trp Leu Glu Pro Thr Tyr His Gly Ile
                35                  40                  45
Val Thr Glu Asn Asp Asn Thr Val Leu Leu Asp Pro Pro Leu Ile
                50                  55                  60
Ala Leu Asp Lys Asp Ala Pro Leu Arg Phe Ala Glu Ser Phe Glu
                65                  70                  75
Val Thr Val Thr Lys Glu Gly Glu Ile Cys Gly Phe Lys Ile His
                80                  85                  90
Gly Gln Asn Val Pro Phe Asp Ala Val Val Val Asp Lys Ser Thr
                95                 100                 105
Gly Glu Gly Val Ile Arg Ser Lys Glu Lys Leu Asp Cys Glu Leu
               110                 115                 120
Gln Lys Asp Tyr Ser Phe Thr Ile Gln Ala Tyr Asp Cys Gly Lys
```

```
                125                 130                 135
Gly Pro Asp Gly Thr Asn Val Lys Lys Ser His Lys Ala Thr Val
                140                 145                 150
His Ile Gln Val Asn Asp Val Asn Glu Tyr Ala Pro Val Phe Lys
                155                 160                 165
Glu Lys Ser Tyr Lys Ala Thr Val Ile Glu Gly Lys Gln Tyr Asp
                170                 175                 180
Ser Ile Leu Arg Val Glu Ala Val Asp Ala Asp Cys Ser Pro Gln
                185                 190                 195
Phe Ser Gln Ile Cys Ser Tyr Glu Ile Ile Thr Pro Asp Val Pro
                200                 205                 210
Phe Thr Val Asp Lys Asp Gly Tyr Ile Lys Asn Thr Glu Lys Leu
                215                 220                 225
Asn Tyr Gly Lys Glu His Gln Tyr Lys Leu Thr Val Thr Ala Tyr
                230                 235                 240
Asp Cys Gly Lys Lys Arg Ala Thr Glu Asp Val Leu Val Lys Ile
                245                 250                 255
Ser Ile Lys Pro Thr Cys Thr Pro Gly Trp Gln Gly Trp Asn Asn
                260                 265                 270
Arg Ile Glu Tyr Glu Pro Gly Thr Gly Ala Leu Ala Val Phe Pro
                275                 280                 285
Asn Ile His Leu Glu Thr Cys Asp Glu Pro Val Ala Ser Val Gln
                290                 295                 300
Ala Thr Val Glu Leu Glu Thr Ser His Ile Gly Lys Gly Cys Asp
                305                 310                 315
Arg Asp Thr Tyr Ser Glu Lys Ser Leu His Arg Leu Cys Gly Ala
                320                 325                 330
Ala Ala Gly Thr Ala Glu Leu Leu Pro Ser Pro Ser Gly Ser Leu
                335                 340                 345
Asn Trp Thr Met Gly Leu Pro Thr Asp Asn Gly His Asp Ser Asp
                350                 355                 360
Gln Val Phe Glu Phe Asn Gly Thr Gln Ala Val Arg Ile Pro Asp
                365                 370                 375
Gly Val Val Ser Val Ser Pro Lys Glu Pro Phe Thr Ile Ser Val
                380                 385                 390
Trp Met Arg His Gly Pro Phe Gly Arg Lys Lys Glu Thr Ile Leu
                395                 400                 405
Cys Ser Ser Asp Lys Thr Asp Met Asn Arg His His Tyr Ser Leu
                410                 415                 420
Tyr Val His Gly Cys Arg Leu Ile Phe Leu Phe Arg Gln Asp Pro
                425                 430                 435
Ser Glu Glu Lys Lys Tyr Arg Pro Ala Glu Phe His Trp Lys Leu
                440                 445                 450
Asn Gln Val Cys Asp Glu Glu Trp His His Tyr Val Leu Asn Val
                455                 460                 465
Glu Phe Pro Ser Val Thr Leu Tyr Val Asp Gly Thr Ser His Glu
                470                 475                 480
Pro Phe Ser Val Thr Glu Asp Tyr Pro Leu His Pro Ser Lys Ile
                485                 490                 495
Glu Thr Gln Leu Val Val Gly Ala Cys Trp Gln Gly Gly Asp Leu
                500                 505                 510
His Met Thr Gln Phe Phe Arg Gly Asn Leu Ala Gly Leu Thr Leu
                515                 520                 525
Arg Ser Gly Lys Leu Ala Asp Lys Lys Val Ile Asp Cys Leu Tyr
                530                 535                 540
Thr Cys Lys Glu Gly Leu Asp Leu Gln Val Leu Glu Asp Ser Gly
```

```
                545                 550                 555
Arg Gly Val Gln Ile Gln Ala His Pro Ser Gln Leu Val Leu Thr
                560                 565                 570
Leu Glu Gly Glu Asp Leu Gly Glu Leu Asp Lys Ala Met Gln His
                575                 580                 585
Ile Ser Tyr Leu Asn Ser Arg Gln Phe Pro Thr Pro Gly Ile Arg
                590                 595                 600
Arg Leu Lys Ile Thr Ser Thr Ile Lys Cys Phe Asn Glu Ala Thr
                605                 610                 615
Cys Ile Ser Val Pro Pro Val Asp Gly Tyr Val Met Val Leu Gln
                620                 625                 630
Pro Glu Glu Pro Lys Ile Ser Leu Ser Gly Val His His Phe Ala
                635                 640                 645
Arg Ala Ala Ser Glu Phe Glu Ser Ser Glu Gly Val Phe Leu Phe
                650                 655                 660
Pro Glu Leu Arg Ile Ile Ser Thr Ile Thr Arg Glu Val Glu Pro
                665                 670                 675
Glu Gly Asp Gly Ala Glu Asp Pro Thr Val Gln Glu Ser Leu Val
                680                 685                 690
Ser Glu Glu Ile Val His Asp Leu Asp Thr Cys Glu Val Thr Val
                695                 700                 705
Glu Gly Glu Glu Leu Asn His Glu Gln Glu Ser Leu Glu Val Asp
                710                 715                 720
Met Ala Arg Leu Gln Gln Lys Gly Ile Glu Val Ser Ser Ser Glu
                725                 730                 735
Leu Gly Met Thr Phe Thr Gly Val Asp Thr Met Ala Ser Tyr Glu
                740                 745                 750
Glu Val Leu His Leu Leu Arg Tyr Arg Asn Trp His Ala Arg Ser
                755                 760                 765
Leu Leu Asp Arg Lys Phe Lys Leu Ile Cys Ser Glu Leu Asn Gly
                770                 775                 780
Arg Tyr Ile Ser Asn Glu Phe Lys Val Glu Val Asn Val Ile His
                785                 790                 795
Thr Ala Asn Pro Met Glu His Ala Asn His Met Ala Ala Gln Pro
                800                 805                 810
Gln Phe Val His Pro Glu His Arg Ser Phe Val Asp Leu Ser Gly
                815                 820                 825
His Asn Leu Ala Asn Pro His Pro Phe Ala Val Val Pro Ser Thr
                830                 835                 840
Ala Thr Val Val Ile Val Val Cys Val Ser Phe Leu Val Phe Met
                845                 850                 855
Ile Ile Leu Gly Val Phe Arg Ile Arg Ala Ala His Arg Arg Thr
                860                 865                 870
Met Arg Asp Gln Asp Thr Gly Lys Glu Asn Glu Met Asp Trp Asp
                875                 880                 885
Asp Ser Ala Leu Thr Ile Thr Val Asn Pro Met Glu Thr Tyr Glu
                890                 895                 900
Asp Gln His Ser Ser Glu Glu Glu Glu Glu Glu Glu Glu Glu Glu
                905                 910                 915
Glu Ser Glu Asp Gly Glu Glu Glu Asp Asp Ile Thr Ser Ala Glu
                920                 925                 930
Ser Glu Ser Ser Glu Val Glu Glu Gly Glu Gln Gly Asp Pro Gln
                935                 940                 945
Asn Ala Thr Arg Gln Gln Gln Leu Glu Trp Asp Asp Ser Thr Leu
                950                 955                 960
Ser Tyr
```

<210> 12
<211> 953
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7509488CD1

<400> 12
Met Leu Arg Arg Pro Ala Pro Arg Ala Val Pro Pro Ala Ala Arg
1 5 10 15
Leu Leu Leu Ala Gly Leu Leu Cys Gly Gly Gly Val Trp Ala Ala
20 25 30
Arg Val Asn Lys His Lys Pro Trp Leu Glu Pro Thr Tyr His Gly
35 40 45
Ile Val Thr Glu Asn Asp Asn Thr Val Leu Leu Asp Pro Pro Leu
50 55 60
Ile Ala Leu Asp Lys Asp Ala Pro Leu Arg Phe Ala Gly Glu Ile
65 70 75
Cys Gly Phe Lys Ile His Gly Gln Asn Val Pro Phe Asp Ala Val
80 85 90
Val Val Asp Lys Ser Thr Gly Glu Gly Val Ile Arg Ser Lys Glu
95 100 105
Lys Leu Asp Cys Glu Leu Gln Lys Asp Tyr Ser Phe Thr Ile Gln
110 115 120
Ala Tyr Asp Cys Gly Lys Gly Pro Asp Gly Thr Asn Val Lys Lys
125 130 135
Ser His Lys Ala Thr Val His Ile Gln Val Asn Asp Val Asn Glu
140 145 150
Tyr Ala Pro Val Phe Lys Glu Lys Ser Tyr Lys Ala Thr Val Ile
155 160 165
Glu Gly Lys Gln Tyr Asp Ser Ile Leu Arg Val Glu Ala Val Asp
170 175 180
Ala Asp Cys Ser Pro Gln Phe Ser Gln Ile Cys Ser Tyr Glu Ile
185 190 195
Ile Thr Pro Asp Val Pro Phe Thr Val Asp Lys Asp Gly Tyr Ile
200 205 210
Lys Asn Thr Glu Lys Leu Asn Tyr Gly Lys Glu His Gln Tyr Lys
215 220 225
Leu Thr Val Thr Ala Tyr Asp Cys Gly Lys Lys Arg Ala Thr Glu
230 235 240
Asp Val Leu Val Lys Ile Ser Ile Lys Pro Thr Cys Thr Pro Gly
245 250 255
Trp Gln Gly Trp Asn Asn Arg Ile Glu Tyr Glu Pro Gly Thr Gly
260 265 270
Ala Leu Ala Val Phe Pro Asn Ile His Leu Glu Thr Cys Asp Glu
275 280 285
Pro Val Ala Ser Val Gln Ala Thr Val Glu Leu Glu Thr Ser His
290 295 300
Ile Gly Lys Gly Cys Asp Arg Asp Thr Tyr Ser Glu Lys Ser Leu
305 310 315
His Arg Leu Cys Gly Ala Ala Ala Gly Thr Ala Glu Leu Leu Pro
320 325 330

```
Ser Pro Ser Gly Ser Leu Asn Trp Thr Met Gly Leu Pro Thr Asp
                335                 340                 345
Asn Gly His Asp Ser Asp Gln Val Phe Glu Phe Asn Gly Thr Gln
                350                 355                 360
Ala Val Arg Ile Pro Asp Gly Val Val Ser Val Ser Pro Lys Glu
                365                 370                 375
Pro Phe Thr Ile Ser Val Trp Met Arg His Gly Pro Phe Gly Arg
                380                 385                 390
Lys Lys Glu Thr Ile Leu Cys Ser Ser Asp Lys Thr Asp Met Asn
                395                 400                 405
Arg His His Tyr Ser Leu Tyr Val His Gly Cys Arg Leu Ile Phe
                410                 415                 420
Leu Phe Arg Gln Asp Pro Ser Glu Glu Lys Lys Tyr Arg Pro Ala
                425                 430                 435
Glu Phe His Trp Lys Leu Asn Gln Val Cys Asp Glu Glu Trp His
                440                 445                 450
His Tyr Val Leu Asn Val Glu Phe Pro Ser Val Thr Leu Tyr Val
                455                 460                 465
Asp Gly Thr Ser His Glu Pro Phe Ser Val Thr Glu Asp Tyr Pro
                470                 475                 480
Leu His Pro Ser Lys Ile Glu Thr Gln Leu Val Val Gly Ala Cys
                485                 490                 495
Trp Gln Gly Gly Asp Leu His Met Thr Gln Phe Phe Arg Gly Asn
                500                 505                 510
Leu Ala Gly Leu Thr Leu Arg Ser Gly Lys Leu Ala Asp Lys Lys
                515                 520                 525
Val Ile Asp Cys Leu Tyr Thr Cys Lys Glu Gly Leu Asp Leu Gln
                530                 535                 540
Val Leu Glu Asp Ser Gly Arg Gly Val Gln Ile Gln Ala His Pro
                545                 550                 555
Ser Gln Leu Val Leu Thr Leu Glu Gly Glu Asp Leu Gly Glu Leu
                560                 565                 570
Asp Lys Ala Met Gln His Ile Ser Tyr Leu Asn Ser Arg Gln Phe
                575                 580                 585
Pro Thr Pro Gly Ile Arg Arg Leu Lys Ile Thr Ser Thr Ile Lys
                590                 595                 600
Cys Phe Asn Glu Ala Thr Cys Ile Ser Val Pro Pro Val Asp Gly
                605                 610                 615
Tyr Val Met Val Leu Gln Pro Glu Glu Pro Lys Ile Ser Leu Ser
                620                 625                 630
Gly Val His His Phe Ala Arg Ala Ala Ser Glu Phe Glu Ser Ser
                635                 640                 645
Glu Gly Val Phe Leu Phe Pro Glu Leu Arg Ile Ile Ser Thr Ile
                650                 655                 660
Thr Arg Glu Val Glu Pro Glu Gly Asp Gly Ala Glu Asp Pro Thr
                665                 670                 675
Val Gln Glu Ser Leu Val Ser Glu Glu Ile Val His Asp Leu Asp
                680                 685                 690
Thr Cys Glu Val Thr Val Glu Gly Glu Glu Leu Asn His Glu Gln
                695                 700                 705
Glu Ser Leu Glu Val Asp Met Ala Arg Leu Gln Gln Lys Gly Ile
                710                 715                 720
Glu Val Ser Ser Ser Glu Leu Gly Met Thr Phe Thr Gly Val Asp
                725                 730                 735
Thr Met Ala Ser Tyr Glu Glu Val Leu His Leu Leu Arg Tyr Arg
                740                 745                 750
```

```
Asn Trp His Ala Arg Ser Leu Leu Asp Arg Lys Phe Lys Leu Ile
                755                 760                 765
Cys Ser Glu Leu Asn Gly Arg Tyr Ile Ser Asn Glu Phe Lys Val
                770                 775                 780
Glu Val Asn Val Ile His Thr Ala Asn Pro Met Glu His Ala Asn
                785                 790                 795
His Met Ala Ala Gln Pro Gln Phe Val His Pro Glu His Arg Ser
                800                 805                 810
Phe Val Asp Leu Ser Gly His Asn Leu Ala Asn Pro His Pro Phe
                815                 820                 825
Ala Val Val Pro Ser Thr Ala Thr Val Val Ile Val Val Cys Val
                830                 835                 840
Ser Phe Leu Val Phe Met Ile Ile Leu Gly Val Phe Arg Ile Arg
                845                 850                 855
Ala Ala His Arg Arg Thr Met Arg Asp Gln Asp Thr Gly Lys Glu
                860                 865                 870
Asn Glu Met Asp Trp Asp Asp Ser Ala Leu Thr Ile Thr Val Asn
                875                 880                 885
Pro Met Glu Thr Tyr Glu Asp Gln His Ser Ser Glu Glu Glu Glu
                890                 895                 900
Glu Glu Glu Glu Glu Glu Glu Ser Glu Asp Gly Glu Glu Glu Asp
                905                 910                 915
Asp Ile Thr Ser Ala Glu Ser Glu Ser Ser Glu Val Glu Glu Gly
                920                 925                 930
Glu Gln Gly Asp Pro Gln Asn Ala Thr Arg Gln Gln Gln Leu Glu
                935                 940                 945
Trp Asp Asp Ser Thr Leu Ser Tyr
                950

<210> 13
<211> 582
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7510060CD1

<400> 13
Met Ile Phe Leu Thr Ala Leu Pro Leu Phe Trp Ile Met Ile Ser
  1               5                  10                  15
Ala Ser Arg Gly Gly His Trp Gly Ala Trp Met Pro Ser Ser Ile
                20                  25                  30
Ser Ala Phe Glu Gly Thr Cys Val Ser Ile Pro Cys Arg Phe Asp
                35                  40                  45
Phe Pro Asp Glu Leu Arg Pro Ala Val Val His Gly Val Trp Tyr
                50                  55                  60
Phe Asn Ser Pro Tyr Pro Lys Asn Tyr Pro Pro Val Val Phe Lys
                65                  70                  75
Ser Arg Thr Gln Val Val His Glu Ser Phe Gln Gly Arg Ser Arg
                80                  85                  90
Leu Leu Gly Asp Leu Gly Leu Arg Asn Cys Thr Leu Leu Leu Ser
                95                 100                 105
Asn Val Ser Pro Glu Leu Gly Gly Lys Tyr Tyr Phe Arg Gly Asp
               110                 115                 120
Leu Gly Gly Tyr Asn Gln Tyr Thr Phe Ser Glu His Ser Val Leu
```

```
                125                 130                 135
Asp Ile Val Asn Thr Pro Asn Ile Val Val Pro Pro Glu Val Val
                140                 145                 150
Ala Gly Thr Glu Val Glu Val Ser Cys Met Val Pro Asp Asn Cys
                155                 160                 165
Pro Glu Leu Arg Pro Glu Leu Ser Trp Leu Gly His Glu Gly Leu
                170                 175                 180
Gly Glu Pro Ala Val Leu Gly Arg Leu Arg Glu Asp Glu Gly Thr
                185                 190                 195
Trp Val Gln Val Ser Leu Leu His Phe Val Pro Thr Arg Glu Ala
                200                 205                 210
Asn Gly His Arg Leu Gly Cys Gln Ala Ser Phe Pro Asn Thr Thr
                215                 220                 225
Leu Gln Phe Glu Gly Tyr Ala Ser Met Asp Val Lys Tyr Pro Pro
                230                 235                 240
Val Ile Val Glu Met Asn Ser Ser Val Glu Ala Ile Glu Gly Ser
                245                 250                 255
His Val Ser Leu Leu Cys Gly Ala Asp Ser Asn Pro Pro Pro Leu
                260                 265                 270
Leu Thr Trp Met Arg Asp Gly Thr Val Leu Arg Glu Ala Val Ala
                275                 280                 285
Glu Ser Leu Leu Leu Glu Leu Glu Glu Val Thr Pro Ala Glu Asp
                290                 295                 300
Gly Val Tyr Ala Cys Leu Ala Glu Asn Ala Tyr Gly Gln Asp Asn
                305                 310                 315
Arg Thr Val Gly Leu Ser Val Met Tyr Ala Pro Trp Lys Pro Thr
                320                 325                 330
Val Asn Gly Thr Met Val Ala Val Glu Gly Glu Thr Val Ser Ile
                335                 340                 345
Leu Cys Ser Thr Gln Ser Asn Pro Asp Pro Ile Leu Thr Ile Phe
                350                 355                 360
Lys Glu Lys Gln Ile Leu Ser Thr Val Ile Tyr Glu Ser Glu Leu
                365                 370                 375
Gln Leu Glu Leu Pro Ala Val Ser Pro Glu Asp Asp Gly Glu Tyr
                380                 385                 390
Trp Cys Val Ala Glu Asn Gln Tyr Gly Gln Arg Ala Thr Ala Phe
                395                 400                 405
Asn Leu Ser Val Glu Phe Ala Pro Val Leu Leu Leu Glu Ser His
                410                 415                 420
Cys Ala Ala Ala Arg Asp Thr Val Gln Cys Leu Cys Val Val Lys
                425                 430                 435
Ser Asn Pro Glu Pro Ser Val Ala Phe Glu Leu Pro Ser Arg Asn
                440                 445                 450
Val Thr Val Asn Glu Ser Glu Arg Glu Phe Val Tyr Ser Glu Arg
                455                 460                 465
Ser Gly Leu Val Leu Thr Ser Ile Leu Thr Leu Arg Gly Gln Ala
                470                 475                 480
Gln Ala Pro Pro Arg Val Ile Cys Thr Ala Arg Asn Leu Tyr Gly
                485                 490                 495
Ala Lys Ser Leu Glu Leu Pro Phe Gln Gly Ala His Arg Leu Met
                500                 505                 510
Trp Ala Lys Ile Gly Pro Val Gly Ala Val Val Ala Phe Ala Ile
                515                 520                 525
Leu Ile Ala Ile Val Cys Tyr Ile Thr Gln Thr Arg Arg Lys Lys
                530                 535                 540
Asn Val Thr Glu Ser Pro Ser Phe Ser Ala Gly Asp Asn Pro Pro
```

```
                545                 550                 555
Val Leu Phe Ser Ser Asp Phe Arg Ile Ser Gly Ala Pro Glu Lys
                560                 565                 570
Tyr Glu Ser Lys Glu Val Ser Thr Leu Glu Ser His
                575                 580
```

<210> 14
<211> 188
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7510226CD1

<400> 14

```
Met Met Glu Glu Arg Ala Ala Ala Ala Val Ala Ala Ala Ala Ser
  1               5                  10                  15
Ser Cys Arg Pro Leu Gly Ser Gly Ala Gly Pro Gly Pro Thr Gly
                20                  25                  30
Ala Ala Pro Val Ser Ala Pro Ala Pro Gly Pro Gly Pro Ala Gly
                35                  40                  45
Lys Gly Gly Gly Gly Gly Gly Ser Pro Gly Pro Thr Ala Gly Pro
                50                  55                  60
Glu Pro Leu Ser Leu Pro Gly Ile Leu His Phe Ile Gln His Glu
                65                  70                  75
Trp Ala Arg Phe Glu Ala Glu Lys Ala Arg Trp Glu Ala Glu Arg
                80                  85                  90
Ala Glu Leu Gln Ala Gln Val Ala Phe Leu Gln Gly Glu Arg Lys
                95                 100                 105
Gly Gln Glu Asn Leu Lys Thr Asp Leu Val Arg Arg Ile Lys Met
               110                 115                 120
Leu Glu Tyr Ala Leu Lys Gln Glu Arg Ala Lys Tyr His Lys Leu
               125                 130                 135
Lys Phe Gly Thr Asp Leu Asn Gln Gly Glu Lys Lys Ala Asp Val
               140                 145                 150
Ser Glu Gln Val Ser Asn Gly Pro Val Glu Ser Val Thr Leu Glu
               155                 160                 165
Asn Ser Pro Leu Val Trp Lys Glu Gly Arg Gln Leu Leu Arg His
               170                 175                 180
Pro Val Pro Ala Gly Phe Ser Phe
               185
```

<210> 15
<211> 478
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7510385CD1

<400> 15

```
Met Gly Arg Ala Gly Ala Ala Ala Val Ile Pro Gly Leu Ala Leu
  1               5                  10                  15
Leu Trp Ala Val Gly Leu Gly Ser Ala Ala Pro Ser Pro Pro Arg
```

```
                  20                  25                  30
Leu Arg Leu Ser Phe Gln Glu Leu Gln Ala Trp His Gly Leu Gln
                  35                  40                  45
Thr Phe Ser Leu Glu Arg Thr Cys Cys Tyr Gln Ala Leu Leu Val
                  50                  55                  60
Asp Glu Glu Arg Gly Arg Leu Phe Val Gly Ala Glu Asn His Val
                  65                  70                  75
Ala Ser Leu Asn Leu Asp Asn Ile Ser Lys Arg Ala Lys Lys Leu
                  80                  85                  90
Ala Trp Pro Ala Pro Val Glu Trp Arg Glu Glu Cys Asn Trp Ala
                  95                 100                 105
Gly Lys Asp Ile Gly Thr Glu Cys Met Asn Phe Val Lys Leu Leu
                 110                 115                 120
His Ala Tyr Asn Arg Thr His Leu Leu Ala Cys Gly Thr Gly Ala
                 125                 130                 135
Phe His Pro Thr Cys Ala Phe Val Glu Val Gly His Arg Ala Glu
                 140                 145                 150
Glu Pro Val Leu Arg Leu Asp Pro Gly Arg Ile Glu Asp Gly Lys
                 155                 160                 165
Gly Lys Ser Pro Tyr Asp Pro Arg His Arg Ala Ala Ser Val Leu
                 170                 175                 180
Val Gly Glu Glu Leu Tyr Ser Gly Val Ala Ala Asp Leu Met Gly
                 185                 190                 195
Arg Asp Phe Thr Ile Phe Arg Ser Leu Gly Gln Arg Pro Ser Leu
                 200                 205                 210
Arg Thr Glu Pro His Asp Ser Arg Trp Leu Asn Glu Pro Lys Phe
                 215                 220                 225
Val Lys Val Phe Trp Ile Pro Glu Ser Glu Asn Pro Asp Asp Asp
                 230                 235                 240
Lys Ile Tyr Phe Phe Phe Arg Glu Thr Ala Val Glu Ala Ala Pro
                 245                 250                 255
Ala Leu Gly Arg Leu Ser Val Ser Arg Val Gly Gln Ile Cys Arg
                 260                 265                 270
Asn Asp Val Gly Gly Gln Arg Ser Leu Val Asn Lys Trp Thr Thr
                 275                 280                 285
Phe Leu Lys Ala Arg Leu Val Cys Ser Val Pro Gly Val Glu Gly
                 290                 295                 300
Asp Thr His Phe Asp Gln Leu Arg Glu Cys Gly Ser Gly Tyr Gly
                 305                 310                 315
Val Gly Glu Gly Gly Ser Gly Ala Asp Ser Gly Ser Pro Arg Arg
                 320                 325                 330
Ser Gly Ala Val Arg Pro Thr Pro Ala Arg Pro Leu Pro Arg Arg
                 335                 340                 345
Gly Cys Val Ser Val Val Leu Ala Gly Pro Pro Asp Pro Ala Ala
                 350                 355                 360
Leu Cys Arg Leu Leu His Val Gln Gln His Leu Pro Gly Leu Cys
                 365                 370                 375
Gly Val Arg Val Gln His Glu Arg Arg Ala Pro Gly Leu Leu Gly
                 380                 385                 390
Thr Leu Cys Thr Gln Gly Gly Ala His Ala Pro Val Gly Val Ile
                 395                 400                 405
Pro Gly Ser Arg Pro Leu Pro Ala Ala Arg His Val Pro Gln Gln
                 410                 415                 420
Asp Leu Trp His Leu Gln Phe His Gln Gly Leu Pro Arg Arg Cys
                 425                 430                 435
His Pro Val Cys Ala Glu Pro Pro Pro His Val Gln Leu Cys Pro
```

```
                440                 445                 450
Ala His Trp Gly Ala Pro Ser Phe Pro Thr Ser Trp Ser Gln Leu
                455                 460                 465
His Leu His Ser Asn Cys Arg Gly Pro Gly Cys Ser Arg
                470                 475
```

<210> 16
<211> 412
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7511618CD1

<400> 16

```
Met Ala Glu Ser Gly Glu Ser Gly Gly Pro Pro Gly Ser Gln Asp
  1               5                  10                  15
Ser Ala Ala Gly Ala Glu Gly Ala Gly Ala Pro Ala Ala Ala Ala
                20                  25                  30
Ser Ala Glu Pro Lys Ile Met Lys Val Thr Val Lys Thr Pro Lys
                35                  40                  45
Glu Lys Glu Glu Phe Ala Val Pro Glu Asn Ser Ser Val Gln Gln
                50                  55                  60
Thr Leu Glu Leu Ala Arg Asn Pro Ala Met Met Gln Glu Met Met
                65                  70                  75
Arg Asn Gln Asp Arg Ala Leu Ser Asn Leu Glu Ser Ile Pro Gly
                80                  85                  90
Gly Tyr Asn Ala Leu Arg Arg Met Tyr Thr Asp Ile Gln Glu Pro
                95                 100                 105
Met Leu Ser Ala Ala Gln Glu Gln Phe Gly Gly Asn Pro Phe Ala
               110                 115                 120
Ser Leu Val Ser Asn Thr Ser Ser Gly Glu Gly Ser Gln Pro Ser
               125                 130                 135
Arg Thr Glu Asn Arg Asp Pro Leu Pro Asn Pro Trp Ala Pro Gln
               140                 145                 150
Thr Ser Gln Ser Ser Ser Ala Ser Ser Gly Thr Ala Ser Thr Val
               155                 160                 165
Gly Gly Thr Thr Gly Ser Thr Ala Ser Gly Thr Ser Gly Gln Ser
               170                 175                 180
Thr Thr Ala Pro Asn Leu Val Pro Gly Val Gly Ala Ser Met Phe
               185                 190                 195
Asn Thr Pro Gly Met Gln Ser Leu Leu Gln Gln Ile Thr Glu Asn
               200                 205                 210
Pro Gln Leu Met Gln Asn Met Leu Ser Ala Pro Tyr Met Arg Ser
               215                 220                 225
Met Met Gln Ser Leu Ser Gln Asn Pro Asp Leu Ala Ala Gln Met
               230                 235                 240
Met Leu Asn Asn Pro Leu Phe Ala Gly Asn Pro Gln Leu Gln Glu
               245                 250                 255
Gln Met Arg Gln Gln Leu Pro Thr Phe Leu Gln Gln Met Gln Asn
               260                 265                 270
Pro Asp Thr Leu Ser Ala Met Ser Asn Pro Arg Ala Met Gln Ala
               275                 280                 285
Leu Leu Gln Ile Gln Gln Gly Leu Gln Thr Leu Ala Thr Glu Ala
               290                 295                 300
```

```
Pro Gly Leu Ile Pro Gly Phe Thr Pro Gly Leu Gly Ala Leu Gly
                305                 310                 315
Ser Thr Gly Gly Ser Ser Gly Thr Asn Gly Ser Asn Ala Thr Pro
                320                 325                 330
Ser Glu Asn Thr Ser Pro Thr Ala Gly Thr Thr Glu Pro Gly His
                335                 340                 345
Gln Gln Phe Ile Gln Gln Met Leu Gln Ala Leu Ala Gly Val Asn
                350                 355                 360
Pro Gln Leu Gln Asn Pro Glu Val Arg Phe Gln Gln Gln Leu Glu
                365                 370                 375
Gln Leu Ser Ala Met Gly Phe Leu Asn Arg Glu Ala Asn Leu Gln
                380                 385                 390
Ala Leu Ile Ala Thr Gly Gly Asp Ile Asn Ala Ala Ile Glu Arg
                395                 400                 405
Leu Leu Gly Ser Gln Pro Ser
                410
```

<210> 17
<211> 136
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 6244135CD1

<400> 17

```
Met Met Asp Gln Glu Glu Lys Thr Glu Glu Gly Ser Gly Pro Cys
  1               5                  10                  15
Ala Glu Ala Gly Ser Pro Asp Gln Glu Gly Phe Phe Asn Leu Leu
                20                  25                  30
Ser His Val Gln Gly Asp Arg Met Glu Gly Gln Arg Cys Ser Leu
                35                  40                  45
Gln Ala Gly Pro Gly Gln Thr Thr Lys Ser Gln Ser Asp Pro Thr
                50                  55                  60
Pro Glu Met Asp Ser Leu Met Asp Met Leu Ala Ser Thr Gln Gly
                65                  70                  75
Arg Arg Met Asp Asp Gln Arg Val Thr Val Ser Ser Leu Pro Gly
                80                  85                  90
Phe Gln Pro Val Gly Ser Lys Asp Gly Ala Gln Lys Arg Ala Gly
                95                 100                 105
Thr Leu Ser Pro Gln Pro Leu Leu Thr Pro Gln Asp Pro Thr Ala
                110                115                 120
Leu Gly Phe Arg Arg Asn Ser Ser Pro Gln Pro Pro Thr Gln Ala
                125                130                 135
Pro
```

<210> 18
<211> 198
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7506689CD1

<400> 18
Met Pro Ala Ser Ala Ala Arg Pro Arg Pro Gly Pro Gly Gln Pro
1 5 10 15
Thr Ala Ser Pro Phe Pro Leu Leu Leu Leu Ala Val Leu Ser Gly
20 25 30
Pro Val Ser Gly Arg Val Pro Arg Ser Val Pro Arg Thr Ser Leu
35 40 45
Pro Ile Ser Glu Ala Asp Ser Cys Leu Thr Arg Phe Ala Val Pro
50 55 60
His Thr Tyr Asn Tyr Ser Val Leu Leu Val Asp Pro Ala Ser His
65 70 75
Thr Leu Tyr Val Gly Ala Arg Asp Thr Ile Phe Ala Leu Ser Leu
80 85 90
Pro Phe Ser Gly Glu Arg Pro Arg Arg Ile Asp Trp Met Val Pro
95 100 105
Glu Ala His Arg Gln Asn Cys Arg Lys Lys Gly Lys Lys Glu Gly
110 115 120
Asp Leu Gly Gly Arg Lys Thr Leu Gln Gln Arg Trp Thr Thr Phe
125 130 135
Leu Lys Ala Asp Leu Leu Cys Pro Gly Leu Ser Met Ala Gly Pro
140 145 150
Pro Val Ser Cys Arg Met Leu Leu Cys Phe Asp Leu Ser Leu Gly
155 160 165
Gln Gly Leu Pro Ser Phe Met Ala Ser Phe Leu Pro Ser Gly Arg
170 175 180
Gly Leu Leu Ser Leu Leu Ser Val Pro Ser Asp His Lys Thr Phe
185 190 195
Gly Gln Cys

<210> 19
<211> 813
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7510185CD1

<400> 19
Met Ala Val Gln Arg Ala Ala Ser Pro Arg Arg Pro Pro Ala Pro
1 5 10 15
Leu Trp Pro Arg Leu Leu Leu Pro Leu Leu Leu Leu Leu Leu Pro
20 25 30
Ala Pro Ser Glu Gly Leu Gly His Ser Ala Glu Leu Ala Phe Ala
35 40 45
Val Glu Pro Ser Asp Asp Val Ala Val Pro Gly Gln Pro Ile Val
50 55 60
Leu Asp Cys Arg Val Glu Gly Thr Pro Pro Val Arg Ile Thr Trp
65 70 75
Arg Lys Asn Gly Val Glu Leu Pro Glu Ser Thr His Ser Thr Leu
80 85 90
Leu Ala Asn Gly Ser Leu Met Ile Arg His Phe Arg Leu Glu Pro
95 100 105
Gly Gly Ser Pro Ser Asp Glu Gly Asp Tyr Glu Cys Val Ala Gln
110 115 120

```
Asn Arg Phe Gly Leu Val Val Ser Arg Lys Ala Arg Ile Gln Ala
                125                 130                 135
Ala Thr Met Ser Asp Phe His Val His Pro Gln Ala Thr Val Gly
                140                 145                 150
Glu Glu Gly Gly Val Ala Arg Phe Gln Cys Gln Ile His Gly Leu
                155                 160                 165
Pro Lys Pro Leu Ile Thr Trp Glu Lys Asn Arg Val Pro Ile Asp
                170                 175                 180
Thr Asp Asn Glu Arg Tyr Thr Leu Leu Pro Lys Gly Val Leu Gln
                185                 190                 195
Ile Thr Gly Leu Arg Ala Glu Asp Gly Gly Ile Phe His Cys Val
                200                 205                 210
Ala Ser Asn Ile Ala Ser Ile Arg Ile Ser His Gly Ala Arg Leu
                215                 220                 225
Thr Val Ser Gly Ser Gly Ser Gly Ala Tyr Lys Glu Pro Ala Ile
                230                 235                 240
Leu Val Gly Pro Glu Asn Leu Thr Leu Thr Val His Gln Thr Ala
                245                 250                 255
Val Leu Glu Cys Val Ala Thr Gly Asn Pro Arg Pro Ile Val Ser
                260                 265                 270
Trp Ser Arg Leu Asp Gly Arg Pro Ile Gly Val Glu Gly Ile Gln
                275                 280                 285
Val Leu Gly Thr Gly Asn Leu Ile Ile Ser Asp Val Thr Val Gln
                290                 295                 300
His Ser Gly Val Tyr Val Cys Ala Ala Asn Arg Pro Gly Thr Arg
                305                 310                 315
Val Arg Arg Thr Ala Gln Gly Arg Leu Val Val Gln Ala Pro Ala
                320                 325                 330
Glu Phe Val Gln His Pro Gln Ser Ile Ser Arg Pro Ala Gly Thr
                335                 340                 345
Thr Ala Met Phe Thr Cys Gln Ala Gln Gly Glu Pro Pro Pro His
                350                 355                 360
Val Thr Trp Leu Lys Asn Gly Gln Val Leu Gly Pro Gly Gly His
                365                 370                 375
Val Arg Leu Lys Asn Asn Asn Ser Thr Leu Thr Ile Ser Gly Ile
                380                 385                 390
Gly Pro Glu Asp Glu Ala Ile Tyr Gln Cys Val Ala Glu Asn Ser
                395                 400                 405
Ala Gly Ser Ser Gln Ala Ser Ala Arg Leu Thr Val Leu Trp Ala
                410                 415                 420
Glu Gly Leu Pro Gly Pro Pro Arg Asn Val Arg Ala Val Ser Val
                425                 430                 435
Ser Ser Thr Glu Val Arg Val Ser Trp Ser Glu Pro Leu Ala Asn
                440                 445                 450
Thr Lys Glu Ile Ile Gly Tyr Val Leu His Ile Arg Lys Ala Ala
                455                 460                 465
Asp Pro Pro Glu Leu Glu Tyr Gln Glu Ala Val Ser Lys Ser Thr
                470                 475                 480
Phe Gln His Leu Val Ser Asp Leu Glu Pro Ser Thr Ala Tyr Ser
                485                 490                 495
Phe Tyr Ile Lys Ala Tyr Thr Pro Arg Gly Ala Ser Ser Ala Ser
                500                 505                 510
Val Pro Thr Leu Ala Ser Thr Leu Gly Glu Ala Pro Ala Pro Pro
                515                 520                 525
Pro Leu Ser Val Arg Val Leu Gly Ser Ser Ser Leu Gln Leu Leu
                530                 535                 540
```

```
Trp Glu Pro Trp Pro Arg Leu Ala Gln His Glu Gly Gly Phe Lys
                545                 550                 555
Leu Phe Tyr Arg Pro Ala Ser Lys Thr Ser Phe Thr Gly Pro Ile
                560                 565                 570
Leu Leu Pro Gly Thr Val Ser Ser Tyr Asn Leu Ser Gln Leu Asp
                575                 580                 585
Pro Thr Ala Val Tyr Glu Val Lys Leu Leu Ala Tyr Asn Gln His
                590                 595                 600
Gly Asp Gly Asn Ala Thr Val Arg Phe Val Ser Leu Arg Gly Ala
                605                 610                 615
Ser Glu Arg Thr Ala Leu Ser Pro Pro Cys Asp Cys Arg Lys Glu
                620                 625                 630
Glu Ala Ala Asn Gln Thr Ser Thr Thr Gly Ile Val Ile Gly Ile
                635                 640                 645
His Ile Gly Val Thr Cys Ile Ile Phe Cys Val Leu Phe Leu Leu
                650                 655                 660
Phe Gly Gln Arg Gly Arg Val Leu Leu Cys Lys Asp Val Glu Asn
                665                 670                 675
Gln Leu Ser Pro Pro Gln Gly Pro Arg Ser Gln Arg Asp Pro Gly
                680                 685                 690
Ile Leu Ala Leu Asn Gly Ala Arg Arg Gly Gln Arg Gly Gln Leu
                695                 700                 705
Gly Arg Asp Glu Lys Arg Val Asp Met Lys Glu Leu Glu Gln Leu
                710                 715                 720
Phe Pro Pro Ala Ser Ala Ala Gly Gln Pro Asp Pro Arg Pro Thr
                725                 730                 735
Asp Pro Ala Ala Pro Ala Pro Cys Glu Glu Thr Gln Leu Ser Leu
                740                 745                 750
Leu Pro Leu Gln Gly Cys Gly Leu Met Glu Gly Lys Thr Thr Glu
                755                 760                 765
Ala Lys Thr Thr Glu Ala Thr Ala Pro Cys Ala Gly Leu Ala Ala
                770                 775                 780
Ala Pro Pro Pro Pro Asp Gly Gly Pro Gly Leu Leu Ser Glu Gly
                785                 790                 795
Gln Ala Ser Arg Pro Ala Ala Ala Arg Val Thr Gln Pro Ala His
                800                 805                 810
Ser Glu Gln
```

<210> 20
<211> 165
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 1420867CD1

<400> 20

```
Met Trp Ile Asp Ile Phe Pro Gln Asp Val Pro Ala Pro Pro Pro
  1               5                  10                  15
Val Asp Ile Lys Pro Arg Gln Pro Ile Ser Tyr Glu Leu Arg Val
                20                  25                  30
Val Ile Trp Asn Thr Glu Asp Val Val Leu Asp Asp Glu Asn Pro
                35                  40                  45
Leu Thr Gly Glu Met Ser Ser Asp Ile Tyr Val Lys Ser Trp Val
```

```
                50                  55                  60
Lys Gly Leu Glu His Asp Lys Gln Glu Thr Asp Val His Phe Asn
                65                  70                  75
Ser Leu Thr Gly Glu Gly Asn Phe Asn Trp Arg Phe Val Phe Arg
                80                  85                  90
Phe Asp Tyr Leu Pro Thr Glu Arg Glu Val Ser Val Arg Arg Arg
                95                 100                 105
Ser Gly Pro Phe Ala Leu Glu Glu Ala Glu Phe Arg Gln Pro Ala
               110                 115                 120
Val Leu Val Leu Gln Asp Pro Trp Ser Cys Ser Tyr Gln Thr Trp
               125                 130                 135
Cys Met Gly Pro Gly Ala Pro Ser Ser Ala Leu Cys Ser Trp Pro
               140                 145                 150
Ala Met Gly Pro Gly Arg Gly Ala Ile Cys Phe Ala Ala Ala Ala
               155                 160                 165


<210> 21
<211> 229
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7512289CD1

<400> 21
Met Val Lys Val Thr Phe Asn Ser Ala Leu Ala Gln Lys Glu Ala
  1               5                  10                  15
Lys Lys Asp Glu Pro Lys Ser Gly Glu Glu Ala Leu Ile Ile Pro
                20                  25                  30
Pro Asp Ala Val Ala Val Asp Cys Lys Asp Pro Asp Asp Val Val
                35                  40                  45
Pro Val Gly Gln Arg Arg Ala Trp Cys Trp Cys Met Cys Phe Gly
                50                  55                  60
Leu Ala Phe Met Leu Ala Gly Val Ile Leu Gly Gly Ala Tyr Leu
                65                  70                  75
Tyr Lys Tyr Phe Ala Leu Gln Pro Asp Asp Val Tyr Tyr Cys Gly
                80                  85                  90
Ile Lys Tyr Ile Lys Asp Asp Val Ile Leu Asn Glu Pro Ser Ala
                95                 100                 105
Asp Ala Pro Ala Ala Leu Tyr Gln Thr Ile Glu Glu Asn Ile Lys
               110                 115                 120
Ile Phe Glu Glu Glu Glu Val Glu Phe Ile Ser Val Pro Val Pro
               125                 130                 135
Glu Phe Ala Asp Ser Asp Pro Ala Asn Ile Val His Asp Phe Asn
               140                 145                 150
Lys Ala Gly Thr Tyr Leu Pro Gln Ser Tyr Leu Ile His Glu His
               155                 160                 165
Met Val Ile Thr Asp Arg Ile Glu Asn Ile Asp His Leu Gly Phe
               170                 175                 180
Phe Ile Tyr Arg Leu Cys His Asp Lys Glu Thr Tyr Lys Leu Gln
               185                 190                 195
Arg Arg Glu Thr Ile Lys Gly Ile Gln Lys Arg Glu Ala Ser Asn
               200                 205                 210
Cys Phe Ala Ile Arg His Phe Glu Asn Lys Phe Ala Val Glu Thr
```

215 220 225
Leu Ile Cys Ser

<210> 22
<211> 1163
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7512447CD1

<400> 22
Met Val Trp Cys Leu Gly Leu Ala Val Leu Ser Leu Val Ile Ser
1 5 10 15
Gln Gly Ala Asp Gly Arg Gly Lys Pro Glu Val Val Ser Val Val
20 25 30
Gly Arg Ala Gly Glu Ser Val Val Leu Gly Cys Asp Leu Leu Pro
35 40 45
Pro Ala Gly Arg Pro Pro Leu His Val Ile Glu Trp Leu Arg Phe
50 55 60
Gly Phe Leu Leu Pro Ile Phe Ile Gln Phe Gly Leu Tyr Ser Pro
65 70 75
Arg Ile Asp Pro Asp Tyr Val Gly Arg Val Arg Leu Gln Lys Gly
80 85 90
Ala Ser Leu Gln Ile Glu Gly Leu Arg Val Glu Asp Gln Gly Trp
95 100 105
Tyr Glu Cys Arg Val Phe Phe Leu Asp Gln His Ile Pro Glu Asp
110 115 120
Asp Phe Ala Asn Gly Ser Trp Val Arg Leu Thr Val Asn Ser Pro
125 130 135
Pro Gln Phe Gln Glu Thr Pro Pro Ala Val Leu Glu Val Gln Glu
140 145 150
Leu Glu Pro Val Thr Leu Arg Cys Val Ala Arg Gly Ser Pro Leu
155 160 165
Pro His Val Thr Trp Lys Leu Arg Gly Lys Asp Leu Gly Gln Gly
170 175 180
Gln Gly Gln Val Gln Val Gln Asn Gly Thr Leu Arg Ile Arg Arg
185 190 195
Val Glu Arg Gly Ser Ser Gly Val Tyr Thr Cys Gln Ala Ser Ser
200 205 210
Thr Glu Gly Ser Ala Thr His Ala Thr Gln Leu Leu Val Leu Gly
215 220 225
Pro Pro Val Ile Val Val Pro Pro Lys Asn Ser Thr Val Asn Ala
230 235 240
Ser Gln Asp Val Ser Leu Ala Cys His Ala Glu Ala Tyr Pro Ala
245 250 255
Asn Leu Thr Tyr Ser Trp Phe Gln Asp Asn Ile Asn Val Phe His
260 265 270
Ile Ser Arg Leu Gln Pro Arg Val Arg Ile Leu Val Asp Gly Ser
275 280 285
Leu Arg Leu Leu Ala Thr Gln Pro Asp Asp Ala Gly Cys Tyr Thr
290 295 300
Cys Val Pro Ser Asn Gly Leu Leu His Pro Pro Ser Ala Ser Ala
305 310 315

```
Tyr Leu Thr Val Leu Cys Met Pro Gly Val Ile Arg Cys Pro Val
                320                 325                 330
Arg Ala Asn Pro Pro Leu Leu Phe Val Ser Trp Thr Lys Asp Gly
                335                 340                 345
Lys Ala Leu Gln Leu Asp Lys Phe Pro Gly Trp Ser Gln Gly Thr
                350                 355                 360
Glu Gly Ser Leu Ile Ile Ala Leu Gly Asn Glu Asp Ala Leu Gly
                365                 370                 375
Glu Tyr Ser Cys Thr Pro Tyr Asn Ser Leu Gly Thr Ala Gly Pro
                380                 385                 390
Ser Pro Val Thr Arg Val Leu Leu Lys Ala Pro Pro Ala Phe Ile
                395                 400                 405
Glu Arg Pro Lys Glu Glu Tyr Phe Gln Glu Val Gly Arg Glu Leu
                410                 415                 420
Leu Ile Pro Cys Ser Ala Gln Gly Asp Pro Pro Pro Val Val Ser
                425                 430                 435
Trp Thr Lys Val Gly Arg Gly Leu Gln Gly Gln Ala Gln Val Asp
                440                 445                 450
Ser Asn Ser Ser Leu Ile Leu Arg Pro Leu Thr Lys Glu Ala His
                455                 460                 465
Gly His Trp Glu Cys Ser Ala Ser Asn Ala Val Ala Arg Val Ala
                470                 475                 480
Thr Ser Thr Asn Val Tyr Val Leu Gly Thr Ser Pro His Val Val
                485                 490                 495
Thr Asn Val Ser Val Val Ala Leu Pro Lys Gly Ala Asn Val Ser
                500                 505                 510
Trp Glu Pro Gly Phe Asp Gly Gly Tyr Leu Gln Arg Phe Ser Val
                515                 520                 525
Trp Tyr Thr Pro Leu Ala Lys Arg Pro Asp Arg Met His His Asp
                530                 535                 540
Trp Val Ser Leu Ala Val Pro Val Gly Ala Ala His Leu Leu Val
                545                 550                 555
Pro Gly Leu Gln Pro His Thr Gln Tyr Gln Phe Ser Val Leu Ala
                560                 565                 570
Gln Asn Lys Leu Gly Ser Gly Pro Phe Ser Glu Ile Val Leu Ser
                575                 580                 585
Ala Pro Glu Gly Leu Pro Thr Thr Pro Ala Ala Pro Gly Leu Pro
                590                 595                 600
Pro Thr Glu Ile Pro Pro Pro Leu Ser Pro Pro Arg Gly Leu Val
                605                 610                 615
Ala Val Arg Thr Pro Arg Gly Val Leu Leu His Trp Asp Pro Pro
                620                 625                 630
Glu Leu Val Pro Lys Arg Leu Asp Gly Tyr Val Leu Glu Gly Arg
                635                 640                 645
Gln Gly Ser Gln Gly Trp Glu Val Leu Asp Pro Ala Val Ala Gly
                650                 655                 660
Thr Glu Thr Glu Leu Leu Val Pro Gly Leu Ile Lys Asp Val Leu
                665                 670                 675
Tyr Glu Phe Arg Leu Val Ala Phe Ala Gly Ser Phe Val Ser Asp
                680                 685                 690
Pro Ser Asn Thr Ala Asn Val Ser Thr Ser Gly Leu Glu Val Tyr
                695                 700                 705
Pro Ser Arg Thr Gln Leu Pro Gly Leu Leu Pro Gln Pro Val Leu
                710                 715                 720
Ala Gly Val Val Gly Gly Val Cys Phe Leu Gly Val Ala Val Leu
                725                 730                 735
```

```
Val Ser Ile Leu Ala Gly Cys Leu Leu Asn Arg Arg Arg Ala Ala
                740                 745                 750
Arg Arg Arg Arg Lys Arg Leu Arg Gln Asp Pro Pro Leu Ile Phe
                755                 760                 765
Ser Pro Thr Gly Lys Ser Ala Ala Pro Ser Ala Leu Gly Ser Gly
                770                 775                 780
Ser Pro Asp Ser Val Ala Lys Leu Lys Leu Gln Gly Ser Pro Val
                785                 790                 795
Pro Ser Leu Arg Gln Ser Leu Leu Trp Gly Asp Pro Ala Gly Thr
                800                 805                 810
Pro Ser Pro His Pro Asp Pro Pro Ser Ser Arg Gly Pro Leu Pro
                815                 820                 825
Leu Glu Pro Ile Cys Arg Gly Pro Asp Gly Arg Phe Val Met Gly
                830                 835                 840
Pro Thr Val Ala Ala Pro Gln Glu Arg Ser Gly Arg Glu Gln Ala
                845                 850                 855
Glu Pro Arg Thr Pro Ala Gln Arg Leu Ala Arg Ser Phe Asp Cys
                860                 865                 870
Ser Ser Ser Ser Pro Ser Gly Ala Pro Gln Pro Leu Cys Ile Glu
                875                 880                 885
Asp Ile Ser Pro Val Ala Pro Pro Pro Ala Ala Pro Pro Ser Pro
                890                 895                 900
Leu Pro Gly Pro Gly Pro Leu Leu Gln Tyr Leu Ser Leu Pro Phe
                905                 910                 915
Phe Arg Glu Met Asn Val Asp Gly Asp Trp Pro Pro Leu Glu Glu
                920                 925                 930
Pro Ser Pro Ala Ala Pro Pro Asp Tyr Met Asp Thr Arg Arg Cys
                935                 940                 945
Pro Thr Ser Ser Phe Leu Arg Ser Pro Glu Thr Pro Pro Val Ser
                950                 955                 960
Pro Arg Glu Ser Leu Pro Gly Ala Val Val Gly Ala Gly Ala Thr
                965                 970                 975
Ala Glu Pro Pro Tyr Thr Ala Leu Ala Asp Trp Thr Leu Arg Glu
                980                 985                 990
Arg Leu Leu Pro Gly Leu Leu Pro Ala Ala Pro Arg Gly Ser Leu
                995                1000                1005
Thr Ser Gln Ser Ser Gly Arg Gly Ser Ala Ser Phe Leu Arg Pro
               1010                1015                1020
Pro Ser Thr Ala Pro Ser Ala Gly Gly Ser Tyr Leu Ser Pro Ala
               1025                1030                1035
Pro Gly Asp Thr Ser Ser Trp Ala Ser Gly Pro Glu Arg Trp Pro
               1040                1045                1050
Arg Arg Glu His Val Val Thr Val Ser Lys Arg Arg Asn Thr Ser
               1055                1060                1065
Val Asp Glu Asn Tyr Glu Trp Asp Ser Glu Phe Pro Gly Asp Met
               1070                1075                1080
Glu Leu Leu Glu Thr Leu His Leu Gly Leu Ala Ser Ser Arg Leu
               1085                1090                1095
Arg Pro Glu Ala Glu Pro Glu Leu Gly Val Lys Thr Pro Glu Glu
               1100                1105                1110
Gly Cys Leu Leu Asn Thr Ala His Val Thr Gly Pro Glu Ala Arg
               1115                1120                1125
Cys Ala Ala Leu Arg Glu Glu Phe Leu Ala Phe Arg Arg Arg Arg
               1130                1135                1140
Asp Ala Thr Arg Ala Arg Leu Pro Ala Tyr Arg Gln Pro Val Pro
               1145                1150                1155
```

His Pro Glu Gln Ala Thr Leu Leu
1160

<210> 23
<211> 2208
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7506758CB1

<400> 23
gaagcttgag aaaatctttc tcttttatat gtaaaacagt gcagaggata tctagcaagt 60
attaaaattt catggaaggg ggcagagatt aagaaaaatg tgaaaaggct ggctgcagct 120
ggcctccgcc gctcatctgc agggcgtgag cgcttggtcc atgcagtgaa gctcttccaa 180
cctgggtcaa cgaaaacgga gaagaaatgg cccaagaaat agatctgagt gctctcaagg 240
agttagaacg cgaggccatt ctccaggtcc tgtaccgaga ccaggcggtt caaaacacag 300
aggaggagag gacacggaaa ctgaaaacac acctgcagca tctccggtgg aaaggagcga 360
agaacacgga ctgggagcac aaagagaagt gctgtgcgcg ctgccagcag gtgctggggt 420
tcctgctgca ccggggcgcc gtgtgccggg gctgcagcca ccgcgtgtgt gcccagtgcc 480
gagtgttcct gaggggggacc catgcctgga agtgcacggt gtgcttcgag gacaggaatg 540
tcaaaataaa aactggagaa tggttctatg aggaacgagc caagaaattt ccaactggag 600
aaggtcagtg caccagatat tctgaaacct ctcaatcaag aggatcccaa atgctctact 660
aaccctattt tgaagcaaca gaatctccca tccagtccgg cacccagtac catattctct 720
ggaggtttta gacacggaag tttaattagc attgacagca cctgtacaga gatgggcaat 780
tttgacaatg ctaatgtcac tggagaaata gaatttgcca ttcattattg cttcaaaacc 840
cattctttag aaatatgcat caaggcctgt aagaaccttg cctatggaga agaaaagaag 900
aaaaagtgca atccgtatgt gaagacctac ctgttgcccg acagatcctc ccagggaaag 960
cgcaagactg gagtccaaag gaacaccgtg gacccgacct ttcaggagac cttgaagtat 1020
caggtggccc ctgcccagct ggtgacccgg cagctgcagg tctcggtgtg gcatctgggc 1080
acgctggccc ggagagtgtt tcttggagaa gtgatcattc ctctggccac gtgggacttt 1140
gaagacagca caacacagtc cttccgctgg catccgctcc gggccaaggc ggagaaatac 1200
gaagacagcg ttcctcagag taatggagag ctcacagtcc gggctaagct ggttctccct 1260
tcacggccca gaaaactcca agaggctcaa gaagggacag atcagccatc acttcatggt 1320
caactttgtt tggtagtgct aggagccaag aatttacctg tgcggccaga tggcaccttg 1380
aactcatttg ttaagggctg tctcactctg ccagaccaac aaaaactgag actgaagtcg 1440
ccagtcctga ggaagcaggc ttgcccccag tggaaacact catttgtctt cagtggcgta 1500
accccagctc agctgaggca gtcaagcttg gagttaactg tctgggatca ggccctcttt 1560
ggaatgaacg accgcttgct tggaggaacc agacttggtt caaagggaga cacagctgtt 1620
ggcggggatg catgctcact atcgaagctc cagtggcaga aagtcctttc cagccccaat 1680
ctatggacag acatgactct tgtcctgcac tgacatgaag gcctcaaggt tccaggttgc 1740
agcaggcgtg aggcactgtg cgtctgcaga ggggctacga accaggtgca gggtcccagc 1800
tggagacccc tttgaccttg agcagtctcc atctgcggcc ctgtcccatg gcttaaccgc 1860
ctattggtat ctgtgtatat ttacgttaaa cacaattatg ttacctaagc ctctggtggg 1920
ttatctcctc tttgagatgt agaaaatggc cagattttaa taaacgttgt tacccatgtc 1980
aaaaaggaga taacccacca gaggcttagg taacataatt gtgtttaacg taaatataca 2040
cagataccaa taggcggtta agccatggga cagggccgca gatggagact gctcaaggtc 2100
aaaggggtct ccagctggga ccctgcacct ggttcgtagc ccctctgcag acgcacagtg 2160
cctcacgcct gctgcaacct ggaaccttga ggccttcatg tcaatgca 2208

<210> 24
<211> 7611
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 4381669CB1

<400> 24
gaaacattgc cattgatgca tctggccact gtgttcagtg gacatggtat gttgtagaag 60
ttatatttct ctagaagaaa ttttcatatg gacagtgttt tcagcatcta tttcagtatg 120
catttccaaa tgagaatact gggacataat gcatatgctc ttcaacaagc tgcctttgtt 180
tgcctacagc tttatagcag gccaaatcga ctcttgtctg gtgtgccact gtacaaaggg 240
gacgtgccaa cccaagataa atctgcagtt acatctttgc tggatggact gaaccaagcc 300
ttcgaggagg tttcatccca gagtggaggg gcaaagaggc agagtataca cttggagcag 360
aagttgtatg atggagtctc agccacctct acttggttgg atgacgttga agaacgttta 420
tttgttgcca cagcactttt accagaagaa acagagactt gtctcttcaa ccaagagatt 480
cttgccaaag acattaagga aatgtctgaa gaaatggata agaacaaaaa cttgttttcc 540
caagcttttc cagagaatgg tgataatcga gatgttattg aagatacttt gggttgtctt 600
ttgggcaggt tatccttgct agactcagta gtgaatcaac gatgtcatca gatgaaagaa 660
agacttcagc aaatactaaa tttccagaat gatctgaaag tgctgtttac atcactggct 720
gacaacaaat acatcattct gcaaaaactg gcaaatgtgt ttgaacagcc cgtagcagaa 780
caaatagagg caatacaaca ggctgaagat ggactcaaag aatttgatgc aggaatcatt 840
gaattaaaga ggcgtggtga caagctacag gtcgagcagc cgtccatgca agaactctcc 900
aagctccagg acatgtatga tgagctgatg atgatcattg gctcccggag gagtggtctg 960
aatcagaacc ttacactcaa gagtcagtat gagagggccc tacaagatct ggctgacctg 1020
ctagaaactg gtcaggagaa gatggcagga gaccagaaaa tcatcgtgtc ttccaaagag 1080
gaaatccagc aactacttga caaacataag gaatactttc agggcctgga atctcatatg 1140
atcttgactg aaacactctt cagaaagata atcagctttg cagtccaaaa ggaaacccag 1200
ttccatacag agctgatggc tcaggcttct gctgtactga aacgggctca caagaggggt 1260
gtggagctgg agtacattct agagacgtgg tcccatctgg atgaggacca gcaggagctc 1320
agcagacagc tggaggtggt ggaaagcagc atcccaagcg tgggtctggt ggaggagaac 1380
gaggacaggc ttattgaccg cataacactc taccagcatt taaaatctag ccttaatgaa 1440
taccagccca aattatatca agtattagat gatgggaaac gacttctgat atccatcagc 1500
tgctcagatc tagaaagcca actaaatcaa cttggagagt gctggctaag taacaccaat 1560
aaaatgtcta aggaacttca cagactggaa acaatattga aacactggac cagatatcaa 1620
agtgaatctg cagatctaat tcactggtta caatctgcaa aagaccggct agaattttgg 1680
actcagcaat ctgtgacagt cccacaagag ctggaaatgg tccgtgatca tctaaatgct 1740
ttcctggagt tttctaaaga agtggatgcc caatcttccc tgaaatcatc tgttctgagt 1800
actggaaatc agctccttcg actaaaaaag gtggacacag ccacgctgcg ctctgagctg 1860
tcgcgcattg atagccagtg gactgacctg ctaaccaata tcccagccgt ccaggagaag 1920
ctccaccagc tccagatgga taaactgcct tcccgccatg ccatttctga agtcatgagt 1980
tggatttctc taatggaaaa tgttattcag aaggatgaag ataatattaa aaattccata 2040
ggttacaagg caattcatga ataccttcag aaatataagg gttttaagat agacattaac 2100
tgtaaacagc tgacagtgga ttttgtgaac cagtccgtgc tacaaatcag cagtcaggat 2160
gtggaaagta agcgtagtga taagactgat tttgctgagc aacttggagc aatgaataaa 2220
agttggcaaa ttctgcaagg tctagtaact gagaagatcc agctgttgga aggcttattg 2280
gaatcttggt cagaatatga aaataatgta caatgtctga aaacatggtt tgaaacccag 2340
gaaaagagac taaaacaaca gcatcgaatt ggagatcagg cttctgttca aaatgcactg 2400
aaagactgtc aggatctgga agatttgatt aaagcaaaag aaaaagaagt agagaaaatt 2460
gagcagaatg gacttgcttt gattcagaac aagaaagaag acgtctctag cattgtcatg 2520
agcacactgc gagagctcgg ccaaacctgg gcaaatttag atcacatggt tggacaatta 2580
aagatactgc tgaaatcagt gcttgaccaa tggagtagtc acaaagtggc ctttgacaag 2640
ataaacagtt acctcatgga ggccagatac tctctttccc gattccgtct gctgactggc 2700
tccttagaag ctgtgcaagt tcaggtggac aatcttcaga atctccaaga tgatctggaa 2760
aaacaggaaa ggagcttaca gaaatttggc tctatcacca accaattatt aaaagagtgt 2820
cacccacccg tgacagaaac tcttaccaat acactgaaag aagtcaacat gagatggaat 2880
aacttgctgg aagagattgc tgagcagcta cagtccagca aggccctact tcagctttgg 2940
caaagataca aggactactc caaacagtgt gcttcgacag ttcagcagca ggaggatcga 3000
accaatgagc tgttgaaggc agccacaaac aaggacattg ccgatgatga ggttgccaca 3060

tggattcaag attgcaacga cctcctcaaa ggactgggca cagttaaaga ttccctcttt 3120
tttctccatg agctgggaga gcaactgaag caacaagtgg atgcttccgc agcatcagct 3180
attcaatcgg atcaactctc tttgagtcaa cacttgtgtg ccctggagca agctctctgc 3240
aaacagcaga cttcattaca ggctggagtt cttgattatg aaacctttgc caagagttta 3300
gaagctttgg aggcctggat agtggaagct gaagaaatac tacaagggca ggaccctagc 3360
cactcatctg acctctccac aatccaggaa aggatggaag aacttaaggg acagatgtta 3420
aaattcagca gcatggctcc agatttagac cgtctaaatg agcttggata taggttaccc 3480
ttgaatgata aggaaatcaa aagaatgcag aatctgaacc gccattggtc tctgatctcc 3540
tctcagacta cagaaagatt cagcaagttg cagtcatttt tgctacaaca tcagactttc 3600
ttggaaaaat gtgaaacatg gatggaattc ctagttcaga cagaacaaaa gttagcagta 3660
gagatttcag gaaattatca gcacctttttg gaacagcaga gagcacacga ggtattgttt 3720
caagccgaga tgttcagtcg tcagcagatt ttgcactcaa tcattattga tgggcaacgt 3780
cttctagaac aaggtcaagt tgatgacagg gatgaattca acctgaaatt gacactcctc 3840
agtaatcaat ggcagggagt gattcgcagg gcccagcaga ggcgggggat cattgacagc 3900
cagattcgcc agtggcagcg ctatagggag atggcagaaa agcttcgtaa atggttggtt 3960
gaagtgtcct acctccccat gagtggtctc ggaagtgttc ctataccact gcaacaagca 4020
aggaccctct ttgatgaagt gcagttcaaa gaaaaagtgt ttctgcggca acaaggcagc 4080
tacatcctga ctgtggaggc tggcaagcaa ctccttctct cggcggacag tggcgctgag 4140
gccgccttgc aggccgaact cgctgaaatc caagagaaat ggaaatcagc cagcatgcgg 4200
ctggaagaac agaagaaaaa actagccttc ttgttgaaag actgggaaaa atgtgagaaa 4260
ggaatagcag attccctgga gaaactacga actttcaaaa agaagctttc gcagtctctc 4320
ccggatcacc atgaagagct ccatgcagaa caaatgcgtt gcaaggaatt agaaaatgca 4380
gttgggagct ggacagatga cttgacccag ttgagcctgc tgaaggacac cctctctgcc 4440
tatatcagtg ctgatgatat ctccattctt aatgaacgcg tagagcttct gcaaaggcag 4500
tgggaagaac tatgccacca gctctcctta aggcggcagc aaataggtga aagattgaat 4560
gaatgggcag tcttcagtga aaagaacaag gaactctgtg agtggttgac tcaaatggaa 4620
agcaaagttt ctcagaatgg agacattctc attgaagaaa tgatagagaa gctcaagaag 4680
gattatcaag aggaaattgc tattgctcaa gagaacaaaa tacagctcca acaaatggga 4740
gaacgacttg ctaaagccag ccatgaaagc aaagcatctg agattgaata caagctggga 4800
aaggtcaacg accggtggca gcatctcctg gacctcattg cagccagggt gaagaagctg 4860
aaggagaccc tggtagccgt gcagcagctt gataagaaca tgagcagcct gaggacctgg 4920
ctcgctcaca tcgagtcaga gctggccaag ccaatagtct acgattcctg taactcggaa 4980
gaaatacaga gaaagcttaa tgagcagcag gagcttcaga gagacataga gaagcacagt 5040
acaggtgttg catctgtcct caacctgtgt gaagtcctgc tgcacgactg tgacgcctgt 5100
gccactgatg ccgagtgtga ctctatacag caggctacga gaaacctgga ccggcggtgg 5160
agaaacattt gtgctatgtc catggaaagg aggctgaaaa tcgaagagac gtggcgattg 5220
tggcagaaat ttctggatga ctattcacgt tttgaagatt ggctgaagtc ttcagaaagg 5280
acagctgctt ttcccagctc ttctggggtg atctatacag ttgccaagga agaactaaag 5340
aaatttgagg ctttccagcg acaggtccac gagtgcctga cgcagctgga actgatcaac 5400
aagcagtacc gccgcctggc cagggagaac cgcactgatt cagcatgtag cctcaaacag 5460
atggttcacg aaggcaacca gagatgggac aacctgcaaa agcgtgtcac ctccatcttg 5520
cgcagactca agcattttat tggccagcgt gaggagtttg agactgcgcg ggacagcatt 5580
ctggtctggc tcacagagat ggatctgcag ctcactaata ttgaacattt ttctgagtgt 5640
gatgttcaag ctaaaataaa gcaactcaag gccttccagc aggaaatttc actgaaccac 5700
aataagattg agcagataat tgcccaagga gaacagctga tagaaaagag tgagcccttg 5760
gatgcagcga tcatcgagga ggaactagat gagctccgac ggtactgcca ggaggtcttc 5820
gggcgtgtgg aaagatacca taagaaactg atccgcctgc ctgtatacga tgagcacgac 5880
ctctcagaca gggagctgga gctggaagac tctgcagctc tgtcggacct gcactggcac 5940
gaccgctctg cagacagcct gctttctcca cagccttcct ccaatctctc cctctcgctc 6000
gctcagcccc tccggagcga gcggtcagga cgagacaccc cggctagtgt ggactccatc 6060
cccctggagt gggatcacga ctatgacctc agtcgggacc tggagtctgc aatgtccaga 6120
gctctgccct ctgaggatga agaaggtcag gatgacaaag atttctacct ccggggagct 6180
gttggcttat caggggacca cagtgcccta gagtcacaga tccgacaact gggcaaagcc 6240
ctggatgata gccgttttca gatacagcaa accgaaaata tcattcgcag caaaactccc 6300
acggggccgg agctagacac cagctacaaa ggctacatga aactgctggg cgaatgcagt 6360
agcagtatag actccgtgaa gagactggag cacaaactga aggaggaaga ggagagcctt 6420

cctggctttg ttaacctgca tagtaccgaa acccaaacgg ctggtgtgat tgaccgatgg 6480
gagcttctcc aggcccaggc attgagcaag gagttgagga tgaagcagaa cctccagaag 6540
tggcagcagt ttaactcaga cttgaacagc atctgggcct ggctgggggа cacggaggag 6600
gagttggaac agctccagcg tctggaactc agcactgaca tccagaccat cgagctccag 6660
atcaaaaagc tcaaggagct ccagaaagct gtggaccacc gcaaagccat catcctctcc 6720
atcaatctct gcagccctga gttcacccag gctgacagca aggagagccg ggacctgcag 6780
gatcgcttgt cgcagatgaa tgggcgctgg gaccgagtgt gctctctgct ggaggagtgg 6840
cggggcctgc tgcaggatgc cctgatgcag tgccagggtt tccatgaaat gagccatggt 6900
ttgcttctta tgctggagaa cattgacaga aggaaaaatg aaattgtccc tattgattct 6960
aaccttgatg cagagatact tcaggaccat cacaaacagc ttatgcaaat aaagcatgag 7020
ctgttggaat cccaactcag agtagcctct ttgcaagaca tgtcttgcca actactggtg 7080
aatgctgaag gaacagactg tttagaagcc aaagaaaaag tccatgttat tggaaatcgg 7140
ctcaaacttc tcttgaagga ggtcagtcgt catatcaagg aactggagaa gttattagac 7200
gtgtcaagta gtcagcagga tttgtcttcc tggtcttctg ctgatgaact ggacacctca 7260
gggtctgtga gtcccacatc aggaaggagc accccaaaca gacagaaaac gccacgaggc 7320
aagtgtagtc tctcacagcc tggaccctct gtcagcagtc cacatagcag gtccacaaaa 7380
ggtggctccg attcctccct ttctgagcca gggccaggtc ggtccggccg cggcttcctg 7440
ttcagagtcc tccgagcagc tcttcccctt cagcttctcc tgctcctcct catcgggctt 7500
gcctgccttg taccaatgtc agaggaagac tacagctgtg ccctctccaa caactttgcc 7560
cggtcattcc accccatgct cagatacacg aatggccctc ctccactctg a 7611

<210> 25
<211> 2259
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7503554CB1

<400> 25
tactagcact gactgtggaa tccttaaggg cccattacat ttctgaagaa gaaagctaag 60
atgaaggaca tgccactccg aattcatgtg ctacttggcc tagctatcac tacactagta 120
caagctgtag ataaaaaagt ggattgtcca cggttatgta cgtgtgaaat caggccttgg 180
tttacaccca gatccattta tatggaagca tctacagtgg attgtaatga tttaggtctt 240
ttaactttcc cagccagatt gccagctaac acacagattc ttctcctaca gactaacaat 300
attgcaaaaa ttgaatactc cacagacttt ccagtaaacc ttactggcct ggatttatct 360
caaaacaatt tatcttcagt caccaatatt aatgtaaaaa agatgcctca gctcctttct 420
gtgtacctag aggaaaacaa acttactgaa ctgcctgaaa aatgtctgtc cgaactgagc 480
aacttacaag aactctatat taatcacaac ttgctttcta caatttcacc tggagccttt 540
attggcctac ataatcttct tcgacttcat ctcaattcaa atagattgca gatgatcaac 600
agtaagtggt ttgatgctct tccaaatcta gagattctga tgattgggga aaatccaatt 660
atcagaatca aagacatgaa ctttaagcct cttatcaatc ttcgcagcct ggttatagct 720
ggtataaacc tcacagaaat accagataac gccttggttg gactggaaaa cttagaaagc 780
atctcttttt acgataacag gcttattaaa gtaccccatg ttgctcttca aaaagttgta 840
aatctcaaat ttttggatct aaataaaaat cctattaata gaatacgaag gggtgatttt 900
agcaatatgc tacacttaaa agagttgggg ataaataata tgcctgagct gatttccatc 960
gatagtcttg ctgtggataa cctgccagat ttaagaaaaa tagaagctac taacaaccct 1020
agattgtctt acattcaccc caatgcattt ttcagactcc ccaagctgga atcactcatg 1080
ctgaacagca atgctctcag tgccctgtgt catccgttgg atgaacatga acaaaatcaa 1140
cattcgattc atggagccag attcactgtt ttgcgtggac ccacctgaat tccaaggtca 1200
gaatgttcgg caagtgcatt tcagggacat gatggaaatt tgtctccctc ttatagctcc 1260
tgagagcttt ccttctaatc taaatgtaga agctgggagc tatgtttcct ttcactgtag 1320
agctactgca gaaccacagc ctgaaatcta ctggataaca ccttctggtc aaaaactctt 1380
gcctaatacc ctgacagaca agttctatgt ccattctgag ggaacactag atataaatgg 1440
cgtaactccc aaagaagggg gtttatatac ttgtatagca actaacctag ttggcgctga 1500

cttgaagtct gttatgatca aagtggatgg atcttttcca caagataaca atggctcttt 1560
gaatattaaa ataagagata ttcaggccaa ttcagttttg gtgtcctgga aagcaagttc 1620
taaaattctc aaatctagtg ttaaatggac agcctttgtc aagactgaaa attctcatgc 1680
tgcgcaaagt gctcgaatac catctgatgt caaggtatat aatcttactc atctgaatcc 1740
atcaactgag tataaaattt gtattgatat tcccaccatc tatcagaaaa acagaaaaaa 1800
atgtgtaaat gtcaccacca aaggtttgca ccctgatcaa aaagagtatg aaaagaataa 1860
taccacaaca cttatggcct gtcttggagg ccttctgggg attattggtg tgatatgtct 1920
tatcagctgc ctctctccag aaatgaactg tgatggtgga cacagctatg tgaggaatta 1980
cttacagaaa ccaacctttg cattaggtga gctttatcct cctctgataa atctctggga 2040
agcaggaaaa gaaaaaagta catcactgaa agtaaaagca actgttatag gtttaccaac 2100
aaatatgtcc taaaaaccac caaggaaacc tactccaaaa atgaacaaaa aaaaaaaaag 2160
cgaaagactg cagttgtgct aaaaacaaaa caaaacaaac aaacaaaaaa aaaagtaaaa 2220
aaagattact ttcgagagag aagtttaagc ttcaccaat 2259

<210> 26
<211> 1522
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7506077CB1

<400> 26
gcctcagggt ccctgctgta gacggggcgg gggagagtac gatgggtggg gcgtggtggg 60
tcgtagggcg ctcgagatgg agcccccagc ttccttgatg gatcgcgggg cgcgagtgcc 120
ctagacaagc cggagctggg accggcaatc gggcgttgat ccttgtcacc tgtcgcagac 180
cctcatccct cccgtgggag ccccctttgg acactctatg accctggacc ctcgggggac 240
ctgaacttga tgcgatggga ggctgtgcag gctcgcggcg gcgcttttcg gattccgagg 300
gggaggagac cgtcccggag ccccggctcc ctctgttgga ccatcagggc gcgcattgga 360
agaacgcggt gggcttctgg ctgctgggcc tttgcaacaa cttctcttat gtggtgatgc 420
tgagtgccgc ccacgacatc cttagccaca agaggacatc gggaaaccag agccatgtgg 480
acccaggccc aacgccgatc ccccacaaca gctcatcacg atttgactgc aactctgtct 540
ctacggccct gtcctacctg ggcctcaccc aggccggcct ctcccctcag cagaccctgc 600
tgtccatgct gggtatccct gccctgctgc tggccagcta tttcttgttg ctcacatctc 660
ctgaggccca ggaccctgga ggggaagaag aagcagagag cgcagcccgg cagcccctca 720
taagaaccga ggccccggag tcgaagccag gctccagctc cagcctctcc cttcgggaaa 780
ggtggacagt gttcaagggt ctgctgtggt acattgttcc cttggtcgta gtttactttg 840
ccgagtattt cattaaccag ggacttttg aactcctctt tttctggaac acttccctga 900
gtcacgctca gcaataccgc tggtaccaga tgctgtacca ggctggcgtc tttgcctccc 960
gctcttctct ccgctgctgt cgcatccgtt tcacctgggc cctggccctg ctgcagtgcc 1020
tcaacctggt gttcctgctg gcagacgtgt ggttcggctt tctgccaagc atctacctcg 1080
tcttcctgat cattctgtat gaggggctcc tgggaggcgc agcctacgtg aacaccttcc 1140
acaacatcgc cctggagacc agtgatgagc accgggagtt tgcaatggcg gccacctgca 1200
tctctgacac actggggatc tccctgtcgg ggctcctggc tttgcctctg catgacttcc 1260
tctgccagct ctcctgatac tcgggatcct caggacgcag gtcacattca cctgtgggca 1320
gagggacagg tcagacaccc aggcccaccc cagagaccct ccatgaactg tgctcccagc 1380
cttcccggca ggtctgggag tagggaaggg ctgaagcctt gtttcccttg caggggggcc 1440
agccattgtc tcccacttgg ggagtttctt cctggcatca tgccttctga ataaatgccg 1500
attttatcca tggaaaaaaa aa 1522

<210> 27
<211> 1099
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7506765CB1

<400> 27
cgcctgcttg gcggccgtcg gcgtgctctt ggccatggcg ctcgggctgc tcatcgccgt 60
gccgctgctg ctgcaggcgg cgccccgagg cgccgcgcac tatgagatga tgggcacctg 120
ccgcatgatc tgcgaccctt acactgccgc gcccggcggg gagcccccgg gtgcaaaggc 180
gcagccaccc ggacccagca ccgccgccct ggaagtcatg caggacctca gcgccaaccc 240
tcctcctcct ttcatccagg gacccaaggg cgacccgggg cgaccgggca agccagggcc 300
gcgggggccc cctggagagc cgggcccgcc tggacccagg ggccctccgg gagagaaggg 360
cgactcgggg cggcccgggc tgccagggct gcaactgacg gcgggcacgg ccagcggcgt 420
cggggtggtg ggcggcgggg ccggggtagg tggcgattcc gagggtgaag tgaccagtgc 480
gctgagcgcc accttcagcg gccccaagat cgccttctat gtgggtctca agagccccca 540
cgaaggctat gaggtgctga agttcgatga cgtggtcacc aacctcggca atcactatga 600
ccccaccacg ggcaagttca gctgccaggt gcgcggcatc tacttcttca cctaccacat 660
cctcatgcgc ggcggcgacg gcaccagcat gtgggcggac ctctgcaaga acgggcaggt 720
ccgggccagc gccattgcac aggacgccga ccagaactac gactacgcca gtaacagcgt 780
ggtgctgcac ttggattcag gggacgaagt gtatgtgaag ctggatggcg ggaaggctca 840
cggaggcaat aataacaagt acagcacgtt ctcgggcttt cttctgtacc cggattaggg 900
gcgcgggggg tgcgaggcgg ggtggctgca ggccgcccgg tctccgcccg ggcgcggctc 960
cttggcaaag gccactctcg attcataaca cttcctgaca tctcctttgg aaaagacaaa 1020
tccctgcgtc ctccctgccc cgctcctggc ctcagtgcgt ctgcgaccca ccacgctcag 1080
ggctgtgctc ctggtctcc 1099

<210> 28
<211> 3311
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7490048CB1

<400> 28
ttagcgcaag cagctctgga gcgggcacag gagcgccggc tccaggcgtg ggaggctggg 60
gcaggctgca cagggctctg ccggagggca ggagggcaca atgctggcgg tgtcactcaa 120
gtggcggctg ggcgtggtga ggcggcgtct caaagacgat gggccctact ccaaaggagg 180
caaggacgca ggaggggccg acgtttccct ggcgtgccgc agacagagca ttccagagga 240
gttccgaggg atcactgtgg tggagctgat caagaaagaa ggcagcacgc tgggcctgac 300
tatctcaggt ggcaccgaca aggatggaaa gcccagggtc tccaacctga gacctggggg 360
acttgcagcc aggagtgatc tgctgaacat tggtgactat attcggtctg tgaacgggat 420
ccacctgacc aggctccgcc acgatgagat catcaccctg ctcaagaatg tgggcgagcg 480
cgtggtgctg gaggtggagt atgagctgcc cccgcccgct cctgagaata accccaggat 540
catttcaaag acagtggacg tctccctcta caaggagggc aatagctttg gctttgtcct 600
tagaggaggt gcccatgaag atgggcacaa gtcccgcccg cttgtcctga cctacgtgcg 660
gcccggtggc cctgccgaca gggagggctc cctgaaggtg ggcgacaggc tgctcagtgt 720
cgatggaatc ccgctgcacg gggccagcca tgccaccgcc ctggccaccc tgcggcagtg 780
cagccacgag gcactctttc aggtggagta tgatgtggcc acccctgaca cggtggctaa 840
tgcttcggga cccttgatgg tggaaatagt caagacgcca gggtctgccc tggggatctc 900
gctcaccacc acctccctcc ggaacaagtc agtcattacc atcgaccgca tcaagccagc 960
cagcgtggtg gacaggagcg gagccctgca ccctggagac cacatcctgt ccatcgatgg 1020
caccagcatg gaacactgct cgctgcttga ggccaccaag ctcctggcca gcatttcaga 1080
gaaggtgcgg ctggagatcc tgcctgtgcc ccagagtcag cggccactga ggccctcaga 1140
ggcagtgaaa gtgcagagga gtgagcagct gcaccgctgg gacccctgcg tgccctcctg 1200
ccacagcccc cggcccggcc actgcaggat gcccacctgg gccacacctg ctggccagga 1260

ccaaagccga tccttgtctt caactccctt ttcctcgccg accttgaacc acgccttttc 1320
ctgcaacaac cccagcaccc ttccccgtgg atcccagccc atgagtcctc gaactacaat 1380
gggccggagg aggcaccgaa gaagggaaca caagacctcg ttgtcgctag cctccagcac 1440
ggtggggccg ggcgggcaga ttgtgcacac ggagaccacg gaggtcgtgc tctgtggaga 1500
ccccctcagc ggctttggcc tccagctcca gggcggcatc ttcgccaccg agaccctgtc 1560
ctcccgaccc ctcgtgtgct tcatcgagcc tgacagtccg gctgagaggt gtgggctgct 1620
gcaggtgggg gaccgtgtcc tgtccatcaa tggcattgcc accgaggacg ggactatgga 1680
ggaagccaac cagctcctgc gggacgccgc actggcccac aaggtcgtgc tggaggtgga 1740
gttcgatgtg gcggagtccg tcatcccaag cagtggcacc ttccacgtga agctgcccaa 1800
gaagcgcagc gtggagctgg gcatcaccat cagctcggcc agcaggaaac gaggggagcc 1860
cttgatcatc tccgacatca agaaaggcag cgtggcacac aggacgggca ccctggagcc 1920
aggcgacaag ctactggcca ttgacaatat ccgcctggac aactgcccca tggaggacgc 1980
cgtgcaaatc ctgcggcagt gcgaggacct ggtgaagctg aagatccgga aggacgagga 2040
caactctgat gagctggaga ccacaggtgc cgtcagttac acagtggagc tgaagcgcta 2100
cgggggtccc ctgggcatca ccatttcggg cacggaggaa ccttttgacc ccattgtcat 2160
ctcaggcctc accaagcgtg gcctggctga gaggactggt gccatccacg tgggggaccg 2220
cattctggcc atcaacaacg ttagcctcaa gggccggccg ctgagcgagg ccatccacct 2280
cctgcaggtg gctggagaga ccgtcacact gaagatcaag aagcaactag accgtcccct 2340
cctaccccgc aagtcgggca gcctcagtga gaccagtgat gctgatgagg acccagcaga 2400
tgccctgaaa ggaggcctgc cagcagcccg cttctcgccg gctgtgccca gtgtggacag 2460
tgctgtggag tcttgggaca gctcggccac cgagggtggc tttgggggcc cagggtccta 2520
tacaccacag gcagcagccc ggggcacgac cccccaggag cggaggcctg gctggctgag 2580
gggcagcccc ccacccaccg agccccggag gacgagctat accccaaccc cagctgacga 2640
gagctttcca gaggaggagg aggaggatga ttgggagccg ccaacgagcc cagcccctgg 2700
ccctgcccga gaggagggct tctggcgcat gtttggagaa gctctcgaag acctggagtc 2760
atgtggtcag tcagaactgc tgagggaact ggaggcatcc atcatgacgg gcaccgtgca 2820
gagggtggcc ctcgagggca ggcctggcca ccggccttgg cagaggggcc gggaggtacg 2880
agcctctcct gcagaaatgg aggagctgtt gctgcctaca cccttggaga tgcacaaggt 2940
gaccctgcac aaggacccca tgcggcatga ctttggtttc agcgtctcag atggcctcct 3000
ggaaaaaggt gtctatgtcc acactgtgcg ccctgatggg ccagcccacc gtggaggcct 3060
ccagcccttc gacagggtcc tgcaggtcaa ccacgtccgt acacgggact tcgactgctg 3120
cctggcggtg ccactcctgg ccgaggcggg tgatgtcttg gagctgatca tcagccgcaa 3180
gccgcacacg gcacacagca gccgggcccc ccgatcgcca ggccccagca gtccccggat 3240
gctctgaagt cagcatgtgg gccgaactcc taaggggcca cccaggcaga tcctggctcc 3300
cagacacgta g 3311

<210> 29
<211> 1874
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 2133585CB1

<400> 29
gctagcacgc agtcccttgg tctcttcggt ctcctgccgc ccccgggaag cgcgctgcgc 60
tgccgaggcg agctaagcgc ccgctcgcca tggggagccc cgcacatcgg cccgcgctgc 120
tgctgctgct gccgcctctg ctgctgctgc tgctgctgcg cgtcccgccc agccgcagct 180
tcccaggatc gggagactca ccactagaag acgatgaagt cgggtattca caccctagat 240
ataaagatac cccgtggtgc tcccccatca aggtgaagta tggggatgtg tactgcaggg 300
cccctcaagg aggatactac aaaacagccc tgggaaccag gtgcgacatt cgctgccaga 360
agggctacga gctgcatggc tcttccctac tgatctgcca gtcaaacaaa cgatggtctg 420
acaaggtcat ctgcagacaa aagcgatgtc ctacccttgc catgccagca aatggagggt 480
ttaagtgtgt agatggtgcc tactttaact cccggtgtga gtattattgt tcaccaggat 540
acacgttgaa aggggagcgg accgtcacat gtatggacaa caaggcctgg agcggccggc 600

cagcctcctg tgtggatatg gaacctccta gaatcaagtg cccaagtgtg aaggaacgca 660
ttgcagaacc caacaaactg acagtccggg tgtcctggga gacacccgaa ggaagagaca 720
cagcagatgg aattcttact gatgtcattc taaaaggcct ccccccaggc tccaactttc 780
cagaaggaga ccacaagatc cagtacacag tctatgacag agctgagaat aagggcactt 840
gcaaatttcg agttaaagta agagtcaaac gctgtggcaa actcaatgcc ccagagaatg 900
gttacatgaa gtgctccagc gacggtgata attatggagc cacctgtgag ttctcctgca 960
tcggcggcta tgagctccag ggtagccctg cccgagtatg tcaatccaac ctggcttggt 1020
ctggcacgga gcccacctgt gcagccatga acgtcaatgt gggtgtcaga acggcagctg 1080
cacttctgga tcagttttat gagaaaagga gactcctcat tgtgtccaca cccacagccc 1140
gaaacctcct ttaccggctc cagctaggaa tgctgcagca agcacagtgt ggccttgatc 1200
ttcgacacat caccgtggtg gagctggtgg gtgtgttccc gactctcatt ggcaggatag 1260
gagcaaagat tatgcctcca gccctagcgc tgcagctcag tttcataacc cttacaaagt 1320
tggagccatg gattttagtc tctgaataat gatgtattaa agaagcagaa ttcaacaacc 1380
aaaaggctgt tgctgcgaat cccactctac tccttcagta tggtgctagt ggataagcat 1440
ggcatggaca aagagcgcta tgtctccctg gtgatgcctg tggccctgtt caacctgatt 1500
gacacttttc ccttgagaaa agaagagatg gtcctacaag ccgaaatgag ccagacctgt 1560
aacacctgac atgatggttc ctctcttggc attcctcttc attgtctaca tagtgacatg 1620
cacacgggaa agccttaaaa atatccttga tgtacagatt ttatttgtat attttaaaag 1680
tctattttat tatgagcttt ctttgcactt aaaaattagc atgctgcttt ttgtacttgg 1740
aagtgtttca taataaatta tatgagccat attattactc tttctaactt tctttactcc 1800
atcatggctg cttgattttg ttagagaaat tagaagccca taaccataca caggctatgc 1860
aacatgttat tcaa 1874

<210> 30
<211> 4211
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7509063CB1

<400> 30
gaaaaatgtg agaaaggaat agcagattcc ctggagaaac tacgaacttt caaaaagaag 60
ctttcgcagt ctctcccgga tcaccatgaa gagctccatg cagaacaaat gcgttgcaag 120
gaattagaaa atgcagttgg gagctggaca gatgacttga cccagttgag cctgctgaag 180
gacaccctct ctgcctatat cagtgctgat gatatctcca ttcttaatga acgcgtagag 240
cttctgcaaa ggcagtggga agaactatgc caccagctct ccttaaggcg gcagcaaata 300
ggtgaaagat tgaatgaatg ggcagtcttc agtgaaaaga acaaggaact ctgtgagtgg 360
ttgactcaaa tggaaagcaa agtttctcag aatggagaca ttctcattga agaaatgata 420
gagaagctca agaaggatta tcaagaggaa attgctattg ctcaagagaa caaaatacag 480
ctccaacaaa tgggagaacg acttgctaaa gccagccatg aaagcaaagc atctgagatt 540
gaatacaagc tgggaaaggt caacgaccgg tggcagcatc tcctggacct cattgcagcc 600
agggtgaaga agctgaagga gaccctggta gccgtgcagc agcttgataa gaacatgagc 660
agcctgagga cctggctcgc tcacatcgag tcagagctgg ccaagccaat agtctacgat 720
tcctgtaact cggaagaaat acagagaaag cttaatgagc agcaggagct tcagagagac 780
atagagaagc acagtacagg tgttgcatct gtcctcaacc tgtgtgaagt cctgctgcac 840
gactgtgacg cctgtgccac tgatgccgag tgtgactcta tacagcaggc tacgagaaac 900
ctggaccggc ggtggagaaa catttgtgct atgtccatgg aaaggaggct gaaaatcgaa 960
gagacgtggc gattgtggca gaaatttctg gatgactatt cacgttttga agattggctg 1020
aagtcttcag aaaggacagc tgcttttccc agctcttctg gggtgatcta tacagttgcc 1080
aaggaagaac taaagaaatt tgaggctttc cagcgacagg tccacgagtg cctgacgcag 1140
ctggaactga tcaacaagca gtaccgccgc ctggccaggg agaaccgcac tgattcagca 1200
tgtagcctca aacagatggt tcacgaaggc aaccagagat gggacaacct gcaaaagcgt 1260
gtcacctcca tcttgcgcag actcaagcat tttattggcc agcgtgagga gtttgagact 1320
gcgcgggaca gcattctggt ctggctcaca gagatggatc tgcagctcac taatattgaa 1380

```
cattttctg agtgtgatgt tcaagctaaa ataaagcaac tcaaggcctt ccagcaggaa 1440
atttcactga accacaataa gattgagcag ataattgccc aaggagaaca gctgatagaa 1500
aagagtgagc ccttggatgc agcgatcatc gaggaggaac tagatgagct ccgacggtac 1560
tgccaggagg tcttcgggcg tgtggaaaga taccataaga aactgatccg cctgcctctc 1620
ccagacgatg agcacgacct ctcagacagg gagctggagc tggaagactc tgcagctctg 1680
tcggacctgc actggcacga ccgctctgca gacagcctgc tttctccaca gccttcctcc 1740
aatctctccc tctcgctcgc tcagcccctc cggagcgagc ggtcaggacg agacaccccg 1800
gctagtgtgg actccatccc cctggagtgg gatcacgact atgacctcag tcgggacctg 1860
gagtctgcaa tgtccagagc tctgccctct gaggatgaag aaggtcagga tgacaaagat 1920
ttctacctcc ggggagctgt tgccttatca gatgtaatga tccccgaaag ccctgaggcc 1980
tatgtaaaac tcacagaaaa tgcaatcaaa aatacctccg gggaccacag tgccctagag 2040
tcacagatcc gacaactggg caaagccctg gatgatagcc gctttcagat acagcaaacc 2100
gaaaatatca ttcgcagcaa aactcccacg gggccggagc tagacaccag ctacaaaggc 2160
tacatgaaac tgctgggcga atgcagtagc agtatagact ccgtgaagag actggagcac 2220
aaactgaagg aggaagagga gagccttcct ggctttgtta acctgcatag taccgaaacc 2280
caaacggctg gtgtgattga ccgatgggag cttctccagg cccaggcatt gagcaaggag 2340
ttgaggatga agcagaacct ccagaagtgg cagcagttta actcagactt gaacagcatc 2400
tgggcctggc tgggggacac ggaggaggag ttggaacagc tccagcgtct ggaactcagc 2460
actgacatcc agaccatcga gctccagatc aaaaagctca aggagctcca gaaagctgtg 2520
gaccaccgca aagccatcat cctctccatc aatctctgca gccctgagtt cacccaggct 2580
gacagcaagg agagccggga cctgcaggat cgcttgtcgc agatgaatgg gcgctgggac 2640
cgagtgtgct ctctgctgga ggagtggcgg ggcctgctgc aggatgccct gatgcagtgc 2700
cagggtttcc atgaaatgag ccatggtttg cttcttatgc tggagaacat tgacagaagg 2760
aaaaatgaaa ttgtccctat tgattctaac cttgatgcag agatacttca ggaccatcac 2820
aaacagctta tgcaaataaa gcatgagctg ttggaatccc aactcagagt agcctctttg 2880
caagacatgt cttgccaact actggtgaat gctgaaggaa cagactgttt agaagccaaa 2940
gaaaaagtcc atgttattgg aaatcggctc aaacttctct tgaaggaggt cagtcgtcat 3000
atcaaggaac tggagaagtt attagacgtg tcaagtagtc agcaggattt gtcttcctgg 3060
tcttctgctg atgaactgga cacctcaggg tctgtgagtc ccacatcagg aaggagcacc 3120
ccaaacagac agaaaacgcc acgaggcaag tgtagtctct cacagcctgg accctctgtc 3180
agcagtccac atagcaggtc cacaaaaggt ggctccgatt cctccctttc tgagccaggg 3240
ccaggtcggt ccggccgcgg cttcctgttc agagtcctcc gagcagctct tcccttcag 3300
cttctcctgc tcctcctcat cgggcttgcc tgccttgtac caatgtcaga ggaagactac 3360
agctgtgccc tctccaacaa ctttgcccgg tcattccacc ccatgctcag atacacgaat 3420
ggccctcctc cactctgaac taagcagatg ccatctgcag aagtgctggt agcataagga 3480
ggatcgggtc ataagcaatc ccaaactacc aacaagagga ccttgatctt ggcgaaagcc 3540
ctcggtgtgg cagctttagc cctcctccag atcacatgtg tgcaaattat ggcttcagag 3600
gtggaagata aacagtgacg ggggaacaaa cagacaacaa gaaggtttgg aagaaatctg 3660
gtttgagact ctgaacctta gcactaagga gattgagtaa ggacctccaa agttccccgg 3720
actcatgaat tctgggccct tggcccattc tgtgcacagc caaggacttc agtagaccat 3780
ctgggcagct ttcccatggt gctgctccaa ccatcagata aatgaccctc ccaagcacca 3840
tgtcagtgtc gtacaatcta ccaaccaacc agtgctgaag agattttaga accttgtaac 3900
atacaatttt taagagctta tatggcagct tccttttac cttgttttcc tttggggcat 3960
gatgttttaa cctttgcttt agaagcacaa gctgtaaatc taaaaggcac tttttttag 4020
aggtataaag aaaaactaga tgtaataaat aagatcatgg aaggctttat gtgaaaaaag 4080
ttgaatgtta tagtaaaaaa aaaaagatat ttatgtatgt acagtttgct aaagccaagt 4140
tttgtttgta ttgatttctt tgcatttatt atagatatta taaaataaaa aaaaaaaaaa 4200
aaaaaaaata t                                                      4211
```

<210> 31
<211> 5006
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature

<223> Incyte ID No: 5496710CB1

<400> 31
gaagaaaaga cggagggagg tgaggaggaa gggaggggga gagacagaga cctagagggg 60
ctgaagaccc agacagagct ggcagagcta ctgagaagag gactggagcg ctctgagagc 120
ctctcaagat cttttggggg agcccaataa atgtgaacat gggatctgtc acgggagctg 180
tcctcaagac gctacttctg ttatctactc aaaattggaa cagagtcgaa gctgggaatt 240
cctatgactg tgatgatcct cttgtgtctg ccttgcctca ggcatccttc agcagttctt 300
ccgagctctc cagcagtcat ggtcctggat ttgcaaggct gaatagaaga gatggagctg 360
gtggctggtc tccacttgtg tctaacaaat accagtggtt gcagattgac cttggagaga 420
gaatggaggt caccgctgtg gccactcaag ggggatatgg tagctccaac tgggtgacca 480
gctacctcct gatgttcagt gatagtggct ggaactggaa acaatatcgc caagaggaca 540
gcatctgggg tttttcagga aatgcaaatg cagacagtgt tgtgtactat agactccagc 600
cttctatcaa agccagattt ctgcgcttca tccctttgga atggaacccc aagggcagaa 660
ttggaatgcg aatcgaagtg ttcggatgtg catacagatc agaagtggtt gatcttgatg 720
gaaaaagttc ccttctctac agatttgatc aaaaatccct gagcccaata aaagacatta 780
tttctttgaa attcaaaacc atgcagagtg atgggattct actccacagg gaagggccaa 840
atggagatca catcacactg caattaagaa gagcaagact ctttttactt attaattcag 900
gtgaagctaa actgccttcc acttccaccc tggtcaatct caccctgggc agcctgctag 960
atgatcagca ttggcattca gtgctcatcc agcgtttggg caaacaagtc aacttcacag 1020
tggacgaaca caggcatcat ttccatgcac ggggagaatt caatctcatg aatcttgatt 1080
atgagatcag ctttggaggg attccagcac ctggaaaatc agtgtcattc ccacatagaa 1140
attttcatgg atgtttagaa aatctctatt ataatggagt ggatatcatt gatttggcca 1200
agcagcaaaa accacagatc attgctatgg gaaatgtgtc attttcttgt tcacaaccac 1260
aatctatgcc cgtgactttt ctgagctcca ggagttattt agcactgcca gacttctctg 1320
gagaggagga ggtttctgcc acttttcaat ttcgaacttg gaataaggca gggcttctgc 1380
tgttcagtga acttcagctg atttcagggg gtatcctcct ctttctgagt gatggaaaac 1440
ttaagtcgaa tctctaccag ccaggaaaat tacccagtga catcacagca ggtgtcgaat 1500
taaatgatgg gcagtggcat tctgtctctt tatctgctaa aaagaatcac ttgagtgtgg 1560
cggtggacgg ccagatggct tctgctgctc ctctgctggg gcctgagcag atttattcgg 1620
gtggcaccta ttattttgga ggttgtcctg acaaaagctt tggatccaaa tgtaaaagtc 1680
cacttggtgg atttcaggga tgtatgaggc tcatttctat cagcggcaaa gtggtagatc 1740
tgatttcagt tcagcagggg tcccttggga acttcagtga ccttcagata gactcatgtg 1800
gcatctcaga caggtgtttg cccaactatt gtgaacacgg tggggagtgt tcccagtcct 1860
ggagcacctt tcattgtaac tgtaccaaca ctggttacag aggagctact tgccataact 1920
ctatctatga gcagtcatgt gaagcctata agcacagagg aaatacttca gggttttact 1980
atatagattc agatggaagt ggtcccctgg aaccatttct tctatattgc aatatgaccg 2040
aaactgcatg gaccatcata cagcacaacg gctctgactt aacaagagtc agaaatacta 2100
atccagagaa cccatatgct gggtttttcg agtatgtggc cagcatggag caacttcagg 2160
ccactattaa ccgtgcagag cactgtgaac aggagtttac ttattactgc aagaagtcac 2220
ggctggtcaa taagcaagat ggaacccctc tgagttggtg ggtaggaaga accaatgaaa 2280
cgcaaaccta ctggggaggt tcttcgcctg atcttcaaaa atgtacttgt ggattagagg 2340
gaaactgcat tgattctcag tattactgca attgtgatgc tgaccggaat gaatggacca 2400
atgacactgg attgcttgct tataaagaac atcttccagt aactaagatc gtgattacag 2460
acacaggccg actgcattca gaagcagctt ataaactggg gcctctgctc tgccggggag 2520
acagatcatt ttggaattca gcttcctttg ataccgaggc ttcatatctt cattttccta 2580
ccttccacgg agaacttagc gcggatgtat ctttcttttt taagacaaca gcttcatctg 2640
gggtattttt agagaacttg gggattgctg attttatacg gatagagctt cgctctccga 2700
cagtagtgac tttttcattt gatgtgggga atgggccttt tgaaatctca gtgcagtcac 2760
ccacccactt caacgacaac cagtggcacc atgtgagggt tgaaaggaac atgaaggagg 2820
cctcccttca agtggatcag ctgacaccaa agacacagcc cgcccccgct gatgggcatg 2880
tcctgttaca gctcaacagt cagctcttcg tgggtggaac ggccaccaga cagagaggct 2940
ttctgggctg cattcggtct ctgcagttga atgggatgac cctggatttg gaagaaagag 3000
cccaggtgac tccagaagtg cagccaggtt gtaggggaca ttgcagcagc tatgggaagt 3060
tatgccgcaa tggagggaaa tgcagagaaa gacccattgg gttcttttgt gactgcactt 3120
tctctgcata cacagggcca ttctgctcaa atgagatttc tgcatatttt ggatctggct 3180

catccgtgat atacaatttt caagaaaatt atcttttaag taaaaactcc agctcccacg 3240
ctgcttcatt tcatggtgat atgaagctga gcagagaaat gatcaaattt agtttccgaa 3300
caacacgaac accaagcttg ctgctttttg tgagctcctt ttacaaagaa tacctttctg 3360
tgatcattgc caaaaatgga agtttgcaga tcaggtacaa gttaaataaa tatcaagagc 3420
ctgatgttgt taactttgat tttaaaaaca tggctgatgg acaacttcac cacataatga 3480
ttaacagaga agaaggagtg gtctttatag agattgacga taatagaagg agacaagttc 3540
acctgtcatc aggcacagaa ttcagtgcag tcaaatctct ggtattgggc aggattttag 3600
aacacagtga tgtggaccag gatactgcac tggcaggtgc gcagggcttc acaggctgcc 3660
tctctgcagt gcagctcagc cacgtggccc ctctgaaggc agctctgcac cccagccacc 3720
cagaccctgt cactgttaca ggacacgtga ctgagtccag ctgtatggcc cagcctggca 3780
ctgatgccac atcaagggaa aggacacact cgtttgcaga tcattctgga acaatagatg 3840
acagagagcc ccttgctaat gcaatcaaaa gtgactctgc agtaattgga ggtctgatag 3900
ctgttgtgat ttttatcttg ctttgcatca ctgccatagc tgttcgcatt tatcagcaga 3960
aaaggttata taaaagaagt gaggcaaaaa ggtcagagaa tgtagacagt gctgaggctg 4020
ttctgaaaag tgagcttaat atacaaaatg cagtcaatga aaatcagaaa gagtacttct 4080
tctgattggc agctatgatt taacataaaa ttatgatagt ttgttttaat agccaggagt 4140
tctcaatgga aaaacgaatg ctcttacact gaatgtacag gcagtgggct tgcagcactg 4200
ccatcttgcc atgtacaggc ttggggtggc tccaggaagc ctcgtccagt gatatatttc 4260
tcatagcatt cattctatgg aacaagaaat tagatattgc tgttaatttt caactgttct 4320
ggtatgatct aaaacaagtt taacctgctt aatggctaca gtttttacat gtgaaaactg 4380
tagccttggt ctcttaacca tgtaatacat aagttttgtt agaggtaaaa attaaatttg 4440
gactataatg tccttgcttt atttgagaac atttgctgtg tttgcttttg tctcctcgtg 4500
gtgttattca tagacctgga aatgcttctc agatcctgtg tggccggtgt ttgcatcttc 4560
agtggccaca agcataataa agccccttg cctttctctg tattatattc aatacaatac 4620
atcaatagtc ttgaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa 4680
aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaagg 4740
ggggggggcgc gcaaagatag gccgccgcgc gcgagggtat ttctgcgggg gcgcacactg 4800
ctgccggggg gggcggagag catacataca cacaactcgt gatacagcgg ggccacaaaa 4860
gggggaggcg ggaagaaaac agcgcgcaaa aaaggggggca tataaaagcg cgcgcacagg 4920
gcgcggtgtt aaaaacgggg ggaggggaga caagcgcggg agacaccaaa tagagcgggg 4980
gaaacccccc gattcagaaa ccggga 5006

<210> 32
<211> 4125
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 72150826CB1

<400> 32
gcgacctccc ggctgcagtc ctacctcttg gaactacccg tgtttccggg cccagccctc 60
gcagcccccc acctcctcgc cccggcccgg ggatccgttg gggccgcgtc ccccacgcgc 120
ccccggagac gccctttccg tgtgcgcccg ggacttggtg aaactttgca ggcgccggct 180
gcgaaatgga tttaatccga ggcgtcttgc tccggctcct gctcctggct tccagcctcg 240
gacccggcgc ggtgtcgctc cgagcggcca tccgaaaacc aggcaaggta gggcctcctc 300
ttgacatcaa gctgggcgca ttgaactgta cggctttcag catccagtgg aaaatgccaa 360
ggcatcctgg aagtcccatc cttgggtaca ctgtctttta ctctgaggtt ggcgcagata 420
aatccctgca ggagcagttg cacagcgtgc ctctcagccg ggacatcccg accacggagg 480
aagtgattgg agatttgaaa ccaggcactg aatatcatgt gagcatagca gcttacagcc 540
aggctggcaa agggcggctg agctctcctc ggcatgtcac cactttgtcc caagattcct 600
gcctgcctcc tgcagctccc cagcagccac atgtcattgt ggtttcggat tctgaggtgg 660
ccctgtcttg gaaacctgga gcgagtgaag gaagcgcccc tattcagtac tattctgtgg 720
aattcatcag gccagatttc gacaagaagt ggacctcaat ccatgagcgg atccagatgg 780
actccatggt tatcaagggc ctcgatccag ataccaacta ccagtttgcc gtgagggcaa 840

tgaattccca tggccccagc ccccgcagct ggcccagtga catcatccgg accctctgcc 900
ctgaggaggc gggaagtggc cgctatggac cccgttatat caccgacatg ggagctggtg 960
aggatgatga aggatttgaa gacgacttag atttggatat ttcctttgag gaggttaaac 1020
cacttcctgc taccaaagga gggaataaga aatttttggt ggaaagcaag aagatgtcta 1080
tatctaaccc aaagaccatt tctaggctca tccccctac ctcagcatct ctccctgtga 1140
ccacggtggc tccccagccc attcccatac agagaaaggg gaagaatggt gtggccataa 1200
tgtcaaggct ctttgacatg ccttgtgatg aaactctctg ctctgctgac agcttctgtg 1260
tcaatgacta cacctggggg ggctcgcgat gccagtgcac cctgggcaaa ggtggtgaga 1320
gctgctcaga agatattgtt atccagtatc ctcagttctt tggccactcc tatgtaacgt 1380
ttgaacctct gaagaattct tatcaggcat ttcaaattac tcttgaattt agggcggagg 1440
cagaggatgg cttactgctc tactgtgggg agaacgaaca cgggaggggg gatttcatgt 1500
ccctggctat catccgacgc tccctgcagt tcaggtttaa ttgtggaact ggggttgcca 1560
tcatcgtaag tgagaccaaa atcaaactag ggggttggca cacggttatg ctctacagag 1620
atgggctgaa cgggctgctg cagctgaaca atggcacccc agtgacaggc cagtctcagg 1680
gccaatacag taaaattact ttccggacac ctctctatct tggtggcgct cccagcgctt 1740
actggttggt tagagcaaca gggacaaacc gaggctttca aggctgtgtg cagtcgctcg 1800
ctgtgaatgg gaggagaatt gacatgaggc cctggcccct gggaaaagca ctcagtgggg 1860
ctgatgtggg ggaatgcagc agtggaatct gtgatgaggc ctcgtgcatc catggtggca 1920
cctgcacagc aatcaaagcc gactcctaca tttgcctctg tccccttggg tttaaaggtc 1980
gacactgtga agatgctttc accttgacca ttcctcagtt cagagagtct ctgagatctt 2040
acgctgcaac tccctggcca ctggagcccc agcattacct ttccttcatg gaatttgaga 2100
tcacatttcg gccagactca ggagatggtg tcctcctgta cagctatgac acaggcagca 2160
aagacttcct gtccatcaac ttggcagggg gccacgtgga gttccgcttt gactgtggct 2220
ctgggaccgg tgtcctcagg agtgaagatc ccctcaccct gggcaactgg cacgagcttc 2280
gtgtatctcg cacagcaaag aatggaatct tacaggtgga taagcagaag atagtggagg 2340
gaatggcaga gggaggcttc acacagatta agtgcaacac agacattttc attggcggag 2400
tccccaatta tgatgatgtg aagaagaact cgggtgtcct gaagcctttc agcgggagca 2460
tccagaagat catcctgaat gaccgaacca tccatgtgaa gcatgacttc acctccggag 2520
tgaatgtgga gaatgcggcc caccctgtg tgagagcccc ttgtgcccat gggggcagct 2580
gccggcccag gaaggagggc tatgactgtg actgcccctt gggctttgag gggcttcact 2640
gccagaaagc gatcatagaa gccattgaga tcccgcagtt tatcggccgc agttacctga 2700
cgtatgacaa cccagatatc ttgaagaggg tgtcaggatc aagatcaaat gtgttcatga 2760
ggtttaaaac aactgccaag gatggccttt tgctgtggag gggagacagc cccatgagac 2820
ccaacagcga cttcatttcc ttgggccttc gggatggagc cctcgtgttc agctataacc 2880
tgggcagtgg tgtggcatcc atcatggtga atggctcctt caacgatggt cggtggcacc 2940
gagttaaggc cgttagggat ggccagtcag gaaagataac cgtggatgac tatggagcca 3000
gaacaggcaa atccccaggc atgatgcggc agcttaacat caatggagct ctgtatgtgg 3060
gtggaatgaa ggaaattgct ctgcacacta acaggcaata tatgagaggg ctcgtgggct 3120
gtatctctca cttcaccctg tccaccgatt accacatttc cctcgtggaa gatgccgtgg 3180
atgggaaaaa catcaacact tgtggagcca agtaacacca gctggccttg tccaagggac 3240
agagccttct attctgagaa tcccaggggc cctcagaccc tgcctgatgc tatatgcaga 3300
ggcccaggga ccaggtgtgt ttcctctcac caagaagaaa gtacacactg atgagaaact 3360
gagaaccaag acaggcatcc ctgggtggcc tttcctgctg acactccacg agctgaccca 3420
gcagaattct ctgtgtagga agcatcggac tttgtccatt gaatatgtag cggctgccag 3480
agatcacaca tcaatgcaaa ttccagagcc tgtctgctat agctcagtga ctgtgttgtg 3540
attcatagta cattaaaaag agagagagag agaaagaatc ccacagggca ctattaaaat 3600
acttctctcc ttccctgact catgacactc ttcctgacag cagaatgact gtgtgacctt 3660
gaacttcaca tttcccacat tggcccttgg attgttcgga ttaacccctt ccactcctca 3720
ctggctggtt cactgtgttc tgactagtcc ataaaaataa agatggaagg agatcaaaca 3780
cagaatcata atcataatgc ggccccctc cccagagctt acctcgaatg tgaaatccct 3840
acttcttata atgcagagtt aaggagctga gcccccatga ttttgtagct gtacttttgt 3900
atgtgtatat ttttatagct gcagattgtc cacacagtat acttttggaa gtttgggtgg 3960
atgactctga attcttgcac acctttccta aaaatttctc ccaagagaga cttgtttggg 4020
ccttttgtgg tgtcagtgga atggataggt aacttgggtg ggcaacttgg ggagctgacc 4080
attctctttt ggtaaatgtt gacggctcaa gaggatgaat atgaa 4125

<210> 33
<211> 3597
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 6799476CB1

<400> 33
cgccgcgtct gagcgcgcgg cccagcgtca cggcggcggc ggcggcggct cctccttgga 60
cccccggagc tccccgcgcc gcggagcagc tggccccagg cccctagagc cccgagagct 120
ccgagagctc cgctcggcgt cccgcgcgcc tccctgccgc tcccgccccg ggctggcgat 180
gctgcgccgc cccgctcccg cgctggcccc ggccgcccgg ctgctgctgg ccgggctgct 240
gtgcggcggc ggggtctggg ccgcgcgagt taacaagcac aagccctggc tggagcccac 300
ctaccacggc atagtcacag agaacgacaa caccgtgctc ctcgaccccc cactgatcgc 360
gctggataaa gatgcgcctc tgcgatttgc agagagtttt gaggtgacag tcaccaaaga 420
aggtgagatt tgtggattta aaattcacgg gcagaatgtc ccctttgatg cagtggtagt 480
ggataaatcc actggtgagg gagtcattcg ctccaaagag aaactggact gtgagctgca 540
gaaagactat tcattcacca tccaggccta tgattgtggg aagggacctg atggcaccaa 600
cgtgaaaaag tctcataaag caactgttca tattcaggtg aacgacgtga atgagtacgc 660
gcccgtgttc aaggagaagt cctacaaagc cacggtcatc gaggggaagc agtacgacag 720
cattttgagg gtggaggccg tggatgccga ctgctcccct cagttcagcc agatttgcag 780
ctacgaaatc atcactccag acgtgccctt tactgttgac aaagatggtt atataaaaaa 840
cacagagaaa ttaaactacg ggaaagaaca tcaatataag ctgaccgtca ctgcctatga 900
ctgtgggaag aaaagagcca cagaagatgt tttggtgaag atcagcatta agcccacctg 960
caccctggg tggcaaggat ggaacaacag gattgagtat gagccgggca ccggcgcgtt 1020
ggccgtcttt ccaaatatcc acctggagac atgtgacgag ccagtcgcct cagtacaggc 1080
cacagtggag ctagaaacca gccacatagg gaaaggctgc gaccgagaca cctactcaga 1140
gaagtccctc caccggctct gtggtgcggc cgcgggcact gccgagctgc tgccatcccc 1200
gagtggatcc ctcaactgga ccatgggcct gcccaccgac aatggccacg acagcgacca 1260
ggtgtttgag ttcaacggca cccaggcagt gaggatcccg gatggcgtcg tgtcggtcag 1320
ccccaaagag ccgttcacca tctcggtgtg gatgagacat gggccattcg gcaggaagaa 1380
ggagacaatt ctttgcagtt ctgataaaac agatatgaat cggcaccact actccctcta 1440
tgtccacggg tgccggctga tcttcctctt ccgtcaggat ccttctgagg agaagaaata 1500
cagacctgca gagttccact ggaagttgaa tcaggtctgt gatgaggaat ggcaccacta 1560
cgtcctcaat gtagaattcc cgagtgtgac tctctatgtg gatggcacgt cccacgagcc 1620
cttctctgtg actgaggatt acccgctcca tccatccaag atagaaactc agctcgtggt 1680
gggggcttgc tggcaaggtg gcgacctgca catgacccag tttttccgag gcaatctggc 1740
tggcttaact ctccgttccg ggaaactcgc ggataagaag gtgatcgact gtctgtatac 1800
ctgcaaggag gggctggacc tgcaggtcct cgaagacagt ggcagaggcg tgcagatcca 1860
agcacacccc agccagttgg tattgacctt ggagggagaa gacctcgggg aattggataa 1920
ggccatgcag cacatctcgt acctgaactc ccggcagttc cccacgcccg gaattcgcag 1980
actcaaaatc accagcacaa tcaagtgttt taacgaggcc acctgcattt cggtcccccc 2040
ggtagatggc tacgtgatgg ttttacagcc cgaggagccc aagatcagcc tgagtggcgt 2100
ccaccatttt gcccgagcag cttctgaatt tgaaagctca gaaggggtgt tccttttccc 2160
tgagcttcgc atcatcagca ccatcacgag agaagtggag cctgaagggg acggggctga 2220
ggaccccaca gttcaagaat cactggtgtc cgaggagatc gtgcacgacc tggatacctg 2280
tgaggtcacg gtggagggag aggagctgaa ccacgagcag gagagcctgg aggtggacat 2340
ggcccgcctg cagcagaagg gcattgaagt gagcagctct gaactgggca tgaccttcac 2400
aggcgtggac accatggcca gctacgagga ggttttgcac ctgctgcgct atcggaactg 2460
gcatgccagg tccttgcttg accggaagtt taagctcatc tgctcagagc tgaatggccg 2520
ctacatcagc aacgaattta aggtggaggt gaatgtaatc cacacggcca accccatgga 2580
acacgccaac cacatggctg cccagccaca gttcgtgcac ccggaacacc gctcctttgt 2640
tgacctgtca ggccacaacc tggccaaccc ccacccgttc gcagtcgtcc ccagcactgc 2700
gacagttgtg atcgtggtgt gcgtcagctt cctggtgttc atgattatcc tgggggtatt 2760

tcggatccgg gccgcacatc ggcggaccat gcgggatcag gacaccggga aggagaacga 2820
gatggactgg gacgactctg ccctgaccat caccgtcaac cccatggaga cctatgagga 2880
ccagcacagc agtgaggagg aggaggaaga ggaagaggaa gaggaaagcg aggacggcga 2940
agaagaggat gacatcacca gcgccgagtc ggagagcagc gaggtggagg agggggagca 3000
gggcgacccc cagaacgcaa cccggcagca gcagctggag tgggatgact ccaccctcag 3060
ctactgaccc gtgccccccgg ccacctcggt ttctgctttc gaagactctg ctgccatccg 3120
ttctcccagt cccaagggtc cacgatgtac aaagtcattt cggccagtag gtgtgcagac 3180
ccctccccgc cacgatcgtc gctgtgcttg gtgtgtagga ccctaggctc cccgcccacc 3240
ctctgcctgg tcgcgctctt cagtcccacg aggagctgac acgtcctctc tggccgccat 3300
ccggctcgca caggggcctc ccagcgcctc aggccccgcg tttgtgtctg gagtctcccc 3360
ccggggagag gacactggcc cctcgcactc cagaaaagcc atgccagctg ggctcgttga 3420
caaagggtaa aacatgctca ctcccacccg gtaatcattt ttttcttttt ttaaaaaaag 3480
tttttatttt ttccaaacta gtgcatgtat aaataatggc aggatggggg gtactgtgta 3540
gatgattaac tgacttttta atattttgta aataaatcgg attccttgtg ccaaaaa 3597

<210> 34
<211> 3569
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7509488CB1

<400> 34
gcgcgtctga gcgcgcggcc agcgtcacgg cggcggcggc ggcggctcct ccttggaccc 60
ccggagctcc ccgcgccgcg gagcagctgg ccccaggccc ctagagcccc ggagagctcc 120
gagagctccg ctcggcgtcc cgcgcgcctc cctgcctgct cccgccccgg gttggcgatg 180
ctgcgccgcc ccgctccccg cgctgtgccc ccggccgccc ggctgctgct ggccgggctg 240
ctgtgcggcg gcggggtctg ggccgcgcga gttaacaagc acaagccctg gctggagccc 300
acctaccacg gcatagtcac agagaacgac aacaccgtgc tcctcgaccc cccactgatc 360
gcgctggata aagatgcgcc tctgcgattt gcaggtgaga tttgtggatt taaaattcac 420
gggcagaatg tccccttga tgcagtggta gtggataaat ccactggtga gggagtcatt 480
cgctccaaag agaaactgga ctgtgagctg cagaaagact attcattcac catccaggcc 540
tatgattgtg ggaagggacc tgatggcacc aacgtgaaaa agtctcataa agcaactgtt 600
catattcagg tgaacgacgt gaatgagtac gcgcccgtgt tcaaggagaa gtcctacaaa 660
gccacggtca tcgaggggaa gcagtacgac agcattttga gggtggaggc cgtggatgcc 720
gactgctccc ctcagttcag ccagatttgc agctacgaaa tcatcactcc agacgtgccc 780
tttactgttg acaaagatgg ttatataaaa aacacagaga aattaaacta cgggaaagaa 840
catcaatata agctgaccgt cactgcctat gactgtggga agaaaagagc cacagaagat 900
gttttggtga agatcagcat taagcccacc tgcacccctg ggtggcaagg atggaacaac 960
aggattgagt atgagccggg caccggcgcg ttggccgtct ttccaaatat ccacctggag 1020
acatgtgacg agccagtcgc ctcagtacag gccacagtgg agctagaaac cagccacata 1080
gggaaaggct gcgaccgaga cacctactca gagaagtccc tccaccggct ctgtggtgcg 1140
gccgcgggca ctgccgagct gctgccatcc ccgagtggat ccctcaactg gaccatgggc 1200
ctgcccaccg acaatggcca cgacagcgac caggtgtttg agttcaacgg cacccaggca 1260
gtgaggatcc cggatggcgt cgtgtcggtc agccccaaag agccgttcac catctcggtg 1320
tggatgagac atgggccatt cggcaggaag aaggagacaa ttctttgcag ttctgataaa 1380
acagatatga atcggcacca ctactccctc tatgtccacg ggtgccggct gatcttcctc 1440
ttccgtcagg atccttctga ggagaagaaa tacagacctg cagagttcca ctggaagttg 1500
aatcaggtct gtgatgagga atggcaccac tacgtcctca atgtagaatt cccgagtgtg 1560
actctctatg tggatggcac gtcccacgag cccttctctg tgactgagga ttacccgctc 1620
catccatcca agatagaaac tcagctcgtg gtgggggctt gctggcaagg tggcgacctg 1680
cacatgaccc agtttttccg aggcaatctg gctggcttaa ctctccgttc cgggaaactc 1740
gcggataaga aggtgatcga ctgtctgtat acctgcaagg aggggctgga cctgcaggtc 1800
ctcgaagaca gtggcagagg cgtgcagatc caagcacacc ccagccagtt ggtattgacc 1860

ttggagggag aagacctcgg ggaattggat aaggccatgc agcacatctc gtacctgaac 1920
tcccggcagt tccccacgcc cggaattcgc agactcaaaa tcaccagcac aatcaagtgt 1980
tttaacgagg ccacctgcat ttcggtcccc ccggtagatg gctacgtgat ggttttacag 2040
cccgaggagc ccaagatcag cctgagtggc gtccaccatt ttgcccgagc agcttctgaa 2100
tttgaaagct cagaaggggt gttccttttc cctgagcttc gcatcatcag caccatcacg 2160
agagaagtgg agcctgaagg ggacggggct gaggacccca cagttcaaga atcactggtg 2220
tccgaggaga tcgtgcacga cctggatacc tgtgaggtca cggtggaggg agaggagctg 2280
aaccacgagc aggagagcct ggaggtggac atggcccgcc tgcagcagaa gggcattgaa 2340
gtgagcagct ctgaactggg catgaccttc acaggcgtgg acaccatggc cagctacgag 2400
gaggttttgc acctgctgcg ctatcggaac tggcatgcca ggtccttgct tgaccggaag 2460
tttaagctca tctgctcaga gctgaatggc cgctacatca gcaacgaatt taaggtggag 2520
gtgaatgtaa tccacacggc caaccccatg gaacacgcca accacatggc tgcccagcca 2580
cagttcgtgc acccggaaca ccgctccttt gttgacctgt caggccacaa cctggccaac 2640
ccccacccgt tcgcagtcgt ccccagcact gcgacagttg tgatcgtggt gtgcgtcagc 2700
ttcctggtgt tcatgattat cctgggggta tttcggatcc gggccgcaca tcggcggacc 2760
atgcgggatc aggacaccgg gaaggagaac gagatggact gggacgactc tgccctgacc 2820
atcaccgtca accccatgga gacctatgag gaccagcaca gcagtgagga ggaggaggaa 2880
gaggaagagg aagaggaaag cgaggacggc gaagaagagg atgacatcac cagcgccgag 2940
tcggagagca gcgaggtgga ggagggggag cagggcgacc cccagaacgc aacccggcag 3000
cagcagctgg agtgggatga ctccaccctc agctactgac ccgtgccccc ggccacctcg 3060
gtttctgctt tcgaagactc tgctgccatc cgttctccca gtcccaaggg tccacgatgt 3120
acaaagtcat ttcggccagt aggtgtgcag acccctcccc gccacgatcg tcgctgtgct 3180
tggtgtgtag gaccctaggc tccccgccca ccctctgcct ggtcgcgctc ttcagtccca 3240
cgaggagctg acacgtcctc tctggccgcc atccggctcg cacaggggcc tcccagcgcc 3300
tcaggccccg cgtttgtgtc tggagtctcc ccccggggag aggacactgg cccctcgcac 3360
tccagaaaag ccatgccagc tgggctcgtt gacaaagggt aaaacatgct cactcccacc 3420
cggtaatcat ttttttcttt ttttaaaaaa agtttttatt ttttccaaac tagtgcatgt 3480
ataaataatg gcaggatggg gggtactgtg tagatgatta actgactttt taatattttg 3540
taaataaatc ggattccttg tgccaaaaa 3569

<210> 35
<211> 2431
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7510060CB1

<400> 35
ttcggctcga ggcgagctgg gctggcggag cagaggtgca gaagcaactg agtccaagtt 60
gtctggcggc ttcaggtgga cccagaagac gtccccaact cagggagatt cagcgatcac 120
tcactcgctg tacagaatga tattcctcac ggcactgcct ctgttctgga ttatgatttc 180
agcctcccga gggggtcact ggggtgcctg gatgccctcg tccatctcgg ccttcgaagg 240
cacgtgcgtc tccatccccc gccgctttga cttcccggat gagctgcggc ccgctgtggt 300
gcatggtgtc tggtacttca atagccccta ccccaagaac tacccccccgg tggtcttcaa 360
gtcgcgcacc caagtagtcc acgagagctt ccagggccgc agccgcctcc tgggggacct 420
gggcctgcga aactgcaccc tcctgctcag caacgtcagc cccgagctgg gcgggaagta 480
ctacttccgt ggggacctgg gcggctacaa ccagtacacc ttctcagagc acagtgtcct 540
ggatatcgtc aacaccccca acatcgtggt gcccccagag gtggtggcag gcacggaggt 600
ggaggtcagc tgcatggtgc cggacaactg cccagagctg cgccctgagc tgagctggct 660
gggccacgag gggctggggg agcccgctgt gctgggccgg ctgcgggagg acgagggcac 720
ctgggtgcag gtgtcactgc tgcacttcgt gcccacgagg gaggccaacg gccacaggct 780
gggctgccag gcctccttcc ccaacaccac cctgcagttc gagggctacg ccagcatgga 840
cgtcaagtac cccccggtga ttgtggagat gaactcctcg gtggaggcca tcgagggctc 900
ccacgtgagc ctgctctgtg gggctgacag caacccccgg ccgctgctga cctggatgcg 960

ggacgggaca gtcctccggg aggcggtggc cgagagcctg ctcctggagc tggaggaggt 1020
gacccccgcc gaagacggcg tctatgcctg cctggccgag aatgcctatg gccaggacaa 1080
ccgcaccgtg gggctcagtg tcatgtatgc accctggaag ccaacagtga acgggacaat 1140
ggtggccgta gaggggggaga cggtctctat cttgtgctcc acacagagca acccggaccc 1200
tattctcacc atcttcaagg agaagcagat cctgtccacg gtcatctacg agagcgagct 1260
gcagctggag ctgccggccg tgtcacccga ggatgatgga gagtactggt gtgtggctga 1320
gaaccagtat ggccagaggg ccactgcctt caacctgtct gtggagttcg cccctgtgct 1380
cctcctggag tcccactgcg cggcagcccg agacacggtg cagtgcctgt gcgtggtgaa 1440
gtccaacccg gagccgtccg tggcctttga gctgccatcg cgcaatgtga ccgtgaacga 1500
gagcgagcgg gagttcgtgt actcggagcg cagcggcctc gtgctcacca gcatcctcac 1560
gctgcggggg caggcccagg ccccgccccg cgtcatctgc accgcgagga acctctatgg 1620
cgccaagagc ctggagctgc ccttccaggg agcccatcga ctgatgtggg ccaagatcgg 1680
gcctgtgggc gccgtggtcg cctttgccat cctgattgcc atcgtctgct acattaccca 1740
gacacgcagg aaaaagaacg tgacagagag ccccagcttc tcggcagggg acaaccctcc 1800
cgtcctgttc agcagcgact tccgcatctc tggggcacca gagaagtacg agtccaaaga 1860
ggtttctacc ctggaatctc actgagtgcc ccaggagagc gagaggcgcc tgggatctga 1920
gaggaggctg ctgggccttc ggggtgagcc cccagagctg gacctgagct attctcactc 1980
ggacctgggg aaacggccca ccaaggacag ctacacgctg acggaggagc tagctgagta 2040
tgctgaaatc cgggtcaagt gaaggagctg ggggcagcct gcgtggctga ccccctcag 2100
gaccctcgct ggcccccact ggctgtgggc tcccttcctc ccaaaagtat cgggggctgg 2160
ggcaggaggg gagtgaggca ggtgacagtg aggtcctggg ggcctgacct ccccctcctt 2220
cccagctgcc cctccctgcc agcacccccа cgccctcatt acggctcctc tctaacctcc 2280
tttaccctca tctgtctgga ggggagctct gtctgtccgt gttatttatt gctacttcct 2340
gcctggtctc ctgcccccac acctggccct ggggcctgta caaaagggac atgaaataaa 2400
tgccccaaag ccaaaaaaaa aaaaaagggg g 2431

<210> 36
<211> 3842
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7510226CB1

<400> 36
ggaaggctgc tccccaccac ctcgccccca ccgctgggga agggttacca tgacaaccaa 60
cctggccggc aatcagagag gatggggcca ctgctggtcg ccctttctct cagagtaaaa 120
gtaattgtcc aggagagtgg tgtttgggtg tggaaggggt aaggctcttt ctctggggcg 180
aagtgggagg agggggggcag aaagcggccc cttcctcagt ctcactccag cgagggggac 240
cgtcgcgcgg ggcctgggtc tccctcgccc tcctcggggc gcgatctccc acccgccggc 300
ccccgcgcgg gcacgctacc cttcggtgta gacccgggtg ggcgaggctc gcgctcgccg 360
ccggcccccc ttcctactgt cccccgccgg gcgcgcgcgc ggtggttgcg gcgggggagg 420
gggggggtgt tggcgcggcc gcgcaggcgc gcgaggcaca gcgggcgcgc gggccggccc 480
ggggcctcca tgatggagga gcgagcggcc gccgcggtcg ccgccgccgc ctcctcctgc 540
cgtccgctcg gctcaggcgc gggccctggc cccactgggg cggccccggt ctccgccccт 600
gcccccgggc cgggcccggc aggtaaggga ggcggcggcg gaggcagccc cgggcccacg 660
gcgggcccgg agcccctgag cctgccgggg atcctgcact ttatccagca cgagtgggcg 720
cgcttcgaag ccgagaaagc ccgctgggag gccgagcgcg ccgagttaca ggctcaggtg 780
gccttccttc agggagagag gaaagggcag gagaatctaa agacggacct ggtgcggcgg 840
atcaagatgc tagagtatgc gctgaagcag gaaagggcca aatatcataa actgaagttt 900
gggacagacc tgaaccaggg ggagaagaaa gcagatgtgt cagaacaagt ctccaatggc 960
cccgtggaat cggtcaccct ggagaacagc ccgttggtgt ggaaggaggg gcggcagctt 1020
ctccgacatc cagtcccagc tggcttcagt ttttaaattg cagaagtaat aagtgctcat 1080
ttgtagaagt acctggaaga ggcgggctac acagacacca tcctcgacat gcggtccaag 1140
cgcgtccgtt ccctgctggg ccgctcgctg gagctcaacg gggcagtgga gccgagtgaa 1200

ggggccccca gggctccacc aggccctgca gggctcagtg gtggggagtc gctgctggtg 1260
aaacagatcg aggagcagat aaagaggaac gcggcaggca aagatggcaa agagcgcttg 1320
ggcggctcag tgctggggca gatccccttc ctgcagaact gcgaggacga agacagcgac 1380
gaggacgatg agctggacag cgtgcagcac aagaagcagc gtgtgaagct cccatccaag 1440
gctctggtgc ccgaaatgga agacgaggat gaggaagacg actctgagga tgctatcaat 1500
gagtttgatt tcctgggctc aggagaggat ggggaagggg ctccagaccc tcggcggtgc 1560
actgtggatg ggagccccca tgagctggaa agccgtcggg tcaaactcca aggcattctg 1620
gctgacctgc gggatgtgga tgggctgccc ccaaaagtga ctggcccgcc tcctggcaca 1680
ccccagcccc ggccacatga agacgtcttc atcatggaca ctatcggggg cggggaggtg 1740
agcctggggg acttggcaga tctcaccgtc accaacgaca acgacctcag ctgcgatctg 1800
tctgacagca aagatgcttt taagaagacg tggaacccca agttcaccct gcgctcgcac 1860
tacgacggca ttcgttccct ggccttccac cacagccagt cggctctgct caccgcctcc 1920
gaggacggca cgctcaagct ctggaacctg cagaaggcgg tcacggccaa gaagaatgcg 1980
gcgctagatg tggaacctat acatgctttc cgggctcaca ggggcccagt gttggctgtg 2040
gctatgggca gcaacagtga atactgctac agtggcgggg cagatgcctg catccatagt 2100
tggaagattc cagacctcag catggatccc tatgatggct acgacccaag cgtgctgagc 2160
cacgtcctgg agggccacgg ggacgccgtg tggggcctgg ccttcagtcc cacctcccag 2220
cgcctggcct cctgttctgc tgatggcacc gtccgcatct gggaccccag cagcagcagc 2280
ccggcctgcc tctgcacctt ccccacagcc agcgagcacg gggtccccac ctcagtggcc 2340
ttcaccagca ccgagcctgc ccacatcgtg gcctccttcc gctctggcga caccgtcttg 2400
tatgacatgg aggttggcag tgccctcctc acgctggagt cccggggcag cagcggtcca 2460
acccagatca accaagtggt gagtcatcca aaccagcctc tcaccatcac cgcccacgac 2520
gacaggggca tccgcttcct ggacaatcgg acaggtaagc cggtgcactc catggttgca 2580
cacctggacg cagtcacctg cctagccgtg gaccccaacg gcgcattcct gatgtcagga 2640
agccatgact gctccctgcg tctctggagc ctggacaaca aaacgtgcgt gcaggagatc 2700
acggcccacc gcaagaagca tgaggaggcc atccacgctg ttgcctgcca ccccagcaag 2760
gccctcattg ccagtgctgg cgctgatgcc ctggccaagg tcttcgtatg atgcccacct 2820
ggccctgccc tggccgccac gctggctggg gtgtagggcc gggcaggtgg ggctgagggg 2880
actcccagcc tcgctggagg tgcagaggcc tggcctctgt ctgtttgccg tggagcctgg 2940
tggtgctaac ggccctttcc ccaggaatcc ctccccaacg gctgccttcc ccaagcatcc 3000
agcctgggga tgggcccctt tcaggaatca cccccgagga tggcctcggg gtttcccccc 3060
gtgccctgac agccgctcct ggtatcagag ctgctttcca gagggcagca cagccagggc 3120
cgggtgtctt tagggatcag tttttccaag actccagaag gtgccaggtt ctccccttga 3180
gctcctgctc ccccatcccc catcctccca gacctgggca ggagctgtgc tgggagaaag 3240
ggtggcagcc gggagcctca gagctgagcc ccggcctggc tgcccctccc ctccggccct 3300
ctccactgct ttccagacac taaccaaggc ttccaggcca gggattgccc aacactccta 3360
gggcagccct cccagcgccc catggggtcg cccatgggta gagacggctt tcttgtgccc 3420
ctccctgggg catggagggt ggagtgggcc tcgggtcccc ctgaactccc tgtatatctg 3480
tataaataac gggattttca tggcgccgcc ccacccgcat tatcactgtg tgatggtctc 3540
agtcagtctc ctccctgtct ccactctttc cctctattta tttcactctc ttgtttggtt 3600
ctaccctgca ccctcggtcc ccttccaggt tcctgtttat aagccccaac ccctctgtcc 3660
ccatcttgta tgtgaaaact tgtctcaata aacccttttgg agtaaaaaaa aaaaaaaaaa 3720
aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa 3780
aaaaaaaaaa caaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa ggcagcgcca 3840
ca 3842

<210> 37
<211> 2892
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7510385CB1

<400> 37

gggcagctca aggctcctcc acacacacac ccgctgaacc ctgagcaccc tgagctgctg 60
agatggggcg ggccggggct gccgccgtga tcccgggcct ggccctgctc tgggcagtgg 120
ggctggggag tgccgccccc agccccccac gccttcggct ctccttccaa gagctccagg 180
cctggcatgg tctccagact ttcagcctgg agcgaacctg ctgctaccag gccttgctgg 240
tggatgagga gcgtggacgc ctgtttgtgg gtgccgagaa ccatgtggcc tccctcaacc 300
tggacaacat cagcaagcgg gccaagaagc tggcctggcc ggcccctgtg gaatggcgag 360
aggagtgcaa ctgggcaggg aaggacattg gtactgagtg catgaacttc gtgaagttgc 420
tgcatgccta caaccgcacc catttgctgg cctgtggcac gggagccttc cacccaacct 480
gtgcctttgt ggaagtgggc caccgggcag aggagcccgt cctccggctg gacccaggaa 540
ggatagagga tggcaagggg aagagtcctt atgaccccag gcatcgggct gcctccgtgc 600
tggtggggga ggagctatac tcaggggtgg cagcagacct catgggacga gactttacca 660
tctttcgcag cctagggcaa cgtccaagtc tccgaacaga gccacacgac tcccgctggc 720
tcaatgagcc caagtttgtc aaggtatttt ggatcccgga gagcgagaac ccagacgacg 780
acaaaatcta cttcttcttt cgtgagacgg cggtagaggc ggcgccggca ctgggacgcc 840
tgtccgtgtc ccgcgttggc cagatctgcc ggaacgacgt gggcggccag cgcagcctgg 900
tcaacaagtg gacgacgttc ctgaaggcgc ggctggtgtg ctcggtgccc ggcgtcgagg 960
gcgacaccca cttcgatcag ctccgtgagt gcgggagtgg gtatggggtt ggggaggggg 1020
gcagcggcgc agactccggg agcccccgcc gcagcggggc tgtgcgccct accccagcta 1080
ggccccttcc ccgcagagga tgtgtttctg ttgtcctcgc gggaccaccg gaccccgctg 1140
ctctatgccg tcttctccac gtccagcagc atcttccagg gctctgcggt gtgcgtgtac 1200
agcatgaacg acgtgcgccg ggccttcttg ggacccttg cacacaagga ggggcccatg 1260
caccagtggg tgtcatacca gggtcgcgtc ccctacccgc ggccaggcat gtgccccagc 1320
aagacctttg gcaccttcag ttccaccaag gacttcccag acgatgtcat ccagtttgcg 1380
cggaaccacc ccctcatgta caactctgtc ctgcccactg gggggcgccc tcttttccta 1440
caagttggag ccaattacac cttcactcaa attgccgcgg accgggttgc agccgctgac 1500
ggacactatg acgtcctctt cattggcaca gacgttggca cggtgctgaa ggtgatctcg 1560
gtccccaagg gcagtaggcc cagcgcagag gggctgctcc tggaggagct gcacgtgttt 1620
gaggactcgg ccgctgtcac cagcatgcaa atttcttcca agaggcacca gctgtacgta 1680
gcctcgcgga gcgcggtggc ccagatcgcg ttgcaccgct gcgctgccca cggccgcgtc 1740
tgcaccgaat gctgtctggc gcgtgacccc tactgcgcct gggacggggt cgcgtgcacg 1800
cgcttccagc ccagtgccaa gaggcggttc cggcggcaag acgtaaggaa tggcgacccc 1860
agcacgttgt gctccggaga ctcgtctcgt cccgcgctgc tggaacacaa ggtgttcggc 1920
gtggagggca gcagcgcctt tctggagtgt gagccccgct cgctgcaggc gcgcgtggag 1980
tggactttcc agcgcgcagg ggtgacagcc cacacccagg tgctggcaga ggagcgcacc 2040
gagcgcaccg cccggggact actgctgcgc aggctgcggc gccgggactc gggcgtgtac 2100
ttgtgcgccg ccgtcgagca gggctttacg caaccgctgc gtcgcctgtc gctgcacgtg 2160
ttgagtgcta cgcaggccga acgactggcg cgggccgagg aggctgcgcc cgccgcgccg 2220
ccgggcccca aactctggta ccgggacttt ctgcagctgg tggagccggg cggaggtggc 2280
agcgcgaact ccctgcgcat gtgccgcccg cagcctgcgc tgcagtcact gcccctggag 2340
tcgcggagaa agggccgtaa ccggaggacc cacgcccctg agcctcgcgc tgagcggggg 2400
ccgcgcagcg caacgcactg gtgaccagac tgtccccacg ccgggaacca agcaggagac 2460
gacaggcgag agaggagcca gacagaccct gaaaagaagg acgggttggg gccgggcaca 2520
ttgggggtca ccggccgatg gagacaccaa ccgacaggcc ctggctgagg gcagctgcgc 2580
gggcttattt attaacagga taacccttga atgtagcagc cccgggaggg cggcacaggt 2640
cgggcgcagg attcagccgg agggaaggga cggggaagcc gagctccaga gcaacgacca 2700
gggccgagga ggtgcctgga gtgcccaccc tgggagacag accccacctc cttgggtagt 2760
gagcagtgag cagaaagctg tgaacaggct gggctgctgg aggtggggcg aggcaggccg 2820
actgtactaa agtaacgcaa taaacgcatt atcagccaaa gctggaatgg ccccagcaga 2880
aaaccccaaa aa 2892

<210> 38
<211> 2756
<212> DNA
<213> Homo sapiens

<220>

<221> misc_feature
<223> Incyte ID No: 7511618CB1

<400> 38
caggcgcgtt acccggcgtg ctccgcgcgg cgccggcgaa gggacgtggg ggaaggggca 60
gggaggagga agcggtggct gctgcggatg tcggtgtgag cgagcggcgc ctgaacacac 120
ggcggctgcc gagcgcctga cccgggcctg cgccagagcc tgcaccgagc tccggggccc 180
cacacccgct acggtggccc tgcgcccgtt gctactgagg cggcgtgctc tgcattcttc 240
gctgtccagg cctgccggct ctggtgtctg ctggctcctc cttgctcgcc tgctccctcc 300
tgcttgcctg agtcaccgcc gccgccgccg ccacagccat ggccgagagt ggtgaaagcg 360
gcggtcctcc gggctcccag gatagcgccg ccggagccga aggtgctggc gccccgcgg 420
ccgctgcctc cgcggagccc aaaatcatga aagtcaccgt gaagaccccg aaggaaaagg 480
aggaattcgc cgtgcccgag aatagctccg tccagcagac gttggaactt gccaggaatc 540
cagcaatgat gcaggagatg atgaggaacc aggaccgagc tttgagcaac ctagaaagca 600
tcccaggggg atataatgct ttaaggcgca tgtacacaga tattcaggaa ccaatgctga 660
gtgctgcaca agagcagttt ggtggtaatc catttgcttc cttggtgagc aatacatcct 720
ctggtgaagg tagtcaacct tcccgtacag aaaatagaga tccactaccc aatccatggg 780
ctccacagac ttcccagagt tcatcagctt ccagcggcac tgccagcact gtgggtggca 840
ctactggtag tactgccagt ggcacttctg ggcagagtac tactgcgcca aatttggtgc 900
ctggagtagg agctagtatg ttcaacacac caggaatgca gagcttgttg caacaaataa 960
ctgaaaaccc acaactgatg caaaacatgt tgtctgcccc ctacatgaga agcatgatgc 1020
agtcactaag ccagaatcct gaccttgctg cacagatgat gctgaataat cccctatttg 1080
ctggaaatcc tcagcttcaa gaacaaatga gacaacagct cccaactttc ctccaacaaa 1140
tgcagaatcc tgatacacta tcagcaatgt caaaccctag agcaatgcag gccttgttac 1200
agattcagca gggtttacag acattagcaa cggaagcccc gggcctcatc ccagggttta 1260
ctcctggctt gggggcatta ggaagcactg gaggctcttc gggaactaat ggatctaacg 1320
ccacacctag tgaaaacaca agtcccacag caggaaccac tgaacctgga catcagcagt 1380
ttattcagca gatgctgcag gctcttgctg gagtaaatcc tcagctacag aatccagaag 1440
tcagatttca gcaacaactg gaacaactca gtgcaatggg atttttgaac cgtgaagcaa 1500
acttgcaagc tctaatagca acaggaggtg atatcaatgc agctattgaa aggttactgg 1560
gctcccagcc atcatagcag catttctgta tcttgaaaaa atgtaattta tttttgataa 1620
cggctcttaa actttaaaat acctgcttta tttcattttg actcttggaa ttctgtgctg 1680
ttataaacaa acccaatatg atgcatttta aggtggagta cagtaagatg tgtgggtttt 1740
tctgtatttt tcttttctgg aacagtggga attaaggcta ctgcatgcat cacttctgca 1800
tttattgtaa ttttttaaaa acatcacctt ttatagttgg gtgaccagat tttgtcctgc 1860
atctgtccag tttatttgct ttttaaacat tagcctatgg tagtaattta tgtagaataa 1920
aagcattaaa aagaagcaaa tcatttgcac tctataattt gtggtacagt attgcttatt 1980
gtgactttgg catgcatttt tgcaaacaat gctgtaagat ttatactact gataattttg 2040
ttttatttgt atacaatata gagtatgcac atttgggact gcatttctgg aaacatactg 2100
caataggctc tctgagcaaa acacctgtaa ctaaaaaagt gaagataaga aaatactctt 2160
aaagctgagt atttcctaat tgtatagaat cttacagcat ctttgacaaa catctcccag 2220
caaaagtgcc ggttagtcag gtttgttgaa aatacagtag aaaagctgat tctggttatc 2280
tctttaagga caattaattg tacagacaca taatgtaaca ttgtctcaac attcattcac 2340
agattgactg taaattacct taatctttgt gcagactgaa ggaacactgt agtatacccc 2400
aaagtgcatt tgcctaggac ttctcagctt ctcccatagg tagtttaaca ggcattaaaa 2460
tttgtaattg aaatgttgct ttcactgaaa aagtgtcttg atgtttcagt tatttttaat 2520
cgccataaaa aaatagaact atcttttggg tttatctgtt ttctcatgca caggcaatac 2580
acaaatttaa aatgagttgt gagccaattg tttctgaagt gttttggtag ttctattaag 2640
aaatagttaa atattgtgct tttcagagcc tcagagaaag ggggacgggg tgggggggtg 2700
gggcagcgga atctgtcctg gatggggcca gcttaaataa tactggcaac caagat 2756

<210> 39
<211> 873
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 6244135CB1

<400> 39
tgggtggctg ggtggggaca cctgcaagtc ctccctaagc cagctaccac cctacactgt 60
tggcctccct tctccaactg tggggatgct gctcaggcct tttgtgacat cacacctgag 120
agtccctggg gtccagtcat tgctgctggg cacagcgagg tccaagctca ggtcgccctg 180
ccccctaccc accatgccag atccagcatc gttgtgggca aacaattatc tggatgatct 240
ttatggggct taagcttggg tgggagcaga tggggcatga gctggggatt tggggatggg 300
gggaatccac acccccacgt cctggacgtt taaaaggccc tctctggcac tgggccgggg 360
cagaggccag cagaaaagtg actggagtcc agggacatga tggatcagga ggagaagacg 420
gaggaaggct caggcccctg tgccgaggcg ggctccccag accaggaggg cttcttcaat 480
ctgctgagcc acgtgcaggg cgaccggatg gagggacagc gctgttcact gcaagccggg 540
ccgggccaga ccaccaagag ccagagcgac cccacccccg agatggacag cctcatggac 600
atgctggcca gtacccaggg ccgccgcatg gatgaccaac gtgtgacagt cagcagcctg 660
cccggcttcc agcccgtggg gtccaaggac ggagcacaga aacgagctgg gaccctcagt 720
ccccaacccc tgctcacccc tcaggacccg accgctctcg gcttccgtcg gaacagcagc 780
ccccagcccc cgacacaagc cccctgaggg cctgaggcat cctgggtctc actcggcccc 840
caaaaactga taaaagaata aaacacttaa atg 873

<210> 40
<211> 2660
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7506689CB1

<400> 40
tgagcagagg ccgtagcttg cgccgcaccc gcggccaggc ggagccaaag atgccggcct 60
ctgctgcgcg gccccgcccg ggtcccgggc agcctacagc ctcgcccttc ccgctactgc 120
tgctggcggt gctgagcggc ccggtatccg gccgcgtccc ccgctcggtg cccagaacct 180
cgcttccaat ctctgaggct gactcctgtc tcacccggtt cgcagtccct cacacataca 240
attactctgt tctccttgtg gatcctgcct cccacacact ttatgttggc gcccgggaca 300
ccatcttcgc tttatccctg cccttctcag gggagagacc ccgcaggatt gactggatgg 360
ttcctgaggc tcacagacag aactgtagga agaaaggcaa gaaagagggg gacctcgggg 420
gccggaagac cctccagcag agatggacga cgtttttgaa agctgacctg ctctgtccag 480
gcctgagcat ggccgggcct ccagtgtcct gcaggatgtt gctgtgcttc gacctgagct 540
tggggcaggg actcccatct tttatggcat cttttcttcc cagtgggagg gggctactat 600
ctctgctgtc tgtgccttcc gaccacaaga cattcggaca gtgctgaatg gtccttcag 660
agaactaaaa catgactgca acagaggact gcctgtcgtg gacaatgatg tgccccagcc 720
cagacctgga gagtgcatca ccaacaacat gaagctccgg cactttggct catctctctc 780
cctgcctgac cgcgtactca ccttcatccg ggaccaccca ctcatggaca ggccagtgtt 840
tccagctgat ggccaccccc tgctggtcac tacagataca gcctatctca gagtcgtggc 900
ccacagggtg accagcctct cagggaaaga gtatgatgtg ctctacctgg ggacagagga 960
tggacacctc caccgagcag tgcggatcgg agctcagctc agcgttcttg aagatctggc 1020
cttattccca gagccacagc cagttgagaa catgaaattg taccacagct ggctcctggt 1080
tggctcccgt actgaggtga cacaagtgaa tacaaccaac tgtggccgtc tccagagctg 1140
ctcagagtgc atcctggccc aggacccagt ctgtgcctgg agcttccggc tggatgagtg 1200
tgtggcccat gccggggagc accgagggtt ggtccaagac atagagtcag cagatgtctc 1260
ctctttgtgt cctaaagagc ctggagaacg tccagtagtg tttgaagttc ccgtggctac 1320
agctgcgcat gtggtcttgc catgttctcc aagctcagca tgggcatcct gtgtgtggca 1380
ccagcccagt ggagtgactg cactcacccc ccggcgggat ggactggagg tggtggtgac 1440
cccaggggcc atgggcgctt atgcctgtga atgtcaggag ggtggggcag cccatgtggt 1500

agcagcttac agcttggtat ggggcagcca gcgagatgct ccgagccggg cccacacagt 1560
gggggcggga ctggctggct tcttcttggg gattctcgca gcatccctga ctctcattct 1620
gattggtcgg cgtcagcagc gacggcgaca gagggaactt ctggctagag acaaggtggg 1680
cctggacctg ggggctccac cttctgggac cacaagctac agccaagacc ctcctcccc 1740
ctctcctgaa gatgagcggt tgccgctggc cctggccaag aggggcagtg gctttggtgg 1800
attctcacca cccttcctgc ttgatccttg cccaagccca gcccacattc ggctaactgg 1860
ggctcctcta gccacatgtg atgaaacatc catctagagc tgggcaaatg accactagtg 1920
tataagtgat cactggaacg gagtgaccac tgagatgctg ggggtcactg ggcctggaag 1980
accatcccag cctctgagtt ctctttgagt atgagtgatt acttggattt tagtatctgt 2040
tctctctgag cctggatggg cttggggcca gacctttgcc tgattcctga ttcccatgag 2100
aaatcagaac tgctttctgc agcaaatcag ggcttccccc taacatctga actcctgtaa 2160
accttcatcc ctggcccccc atcttgggcc cattagtttt ggggatgggg cacagggcat 2220
agctatgact ttgctttctg gttggagcct ggccggaagg aagagccctg gaggtggttg 2280
ggggcaaatg tgccctgagt ccttggggtg gttctgctta ttcttcaagt ttatctgaat 2340
ctgtggggag tgcatgatcc ccatgttgca atatggagtc tctgccctga gatcttcccc 2400
atctcagttt tccttccatg aaagagtacg tgtaaataca tagtgttcat aagagacctg 2460
gccacatgtg catgaatgac tgggccgatg cttaggaata tctctgaggg atggggaggg 2520
agacaattgg aatgggggcg ttgactgcag aactttagac cctgtagcca tagacggcat 2580
ggcgtccttt ctaaacggaa tgaaacactg ccctctcccc atgcccacaa cttactctac 2640
tttgctacca aaactctaaa 2660

<210> 41
<211> 2761
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7510185CB1

<400> 41
atggctgtgc agcgcgccgc gtctccgcgc cgcccgcccg ccccgctctg gccccggctc 60
ctgctgccgc tgctgttgct gctgctgccc gcgccgagcg agggtcttgg ccactctgcc 120
gaactggcat ttgctgtgga gccaagtgat gatgttgccg tccccgggca gcctatagtg 180
ctggactgca gggtggaggg gacccctcca gtgcgaatca cctggaggaa gaatggggta 240
gagctgccag agagtaccca ctccaccttg ctggccaatg ggtccttgat gatccgtcac 300
ttcaggctgg agccgggagg cagcccttcg gatgaaggtg actatgagtg tgtggcccag 360
aaccgctttg ggctggtggt cagccggaag gctcgcatcc aagctgcaac catgtcggac 420
ttccacgtgc atccccaggc caccgtgggt gaggagggtg gtgtggcccg cttccagtgc 480
caaatccatg ggcttcccaa acccctgatc acttgggaga agaacagagt cccaattgac 540
acggacaatg agaggtacac attgctgccc aagggggtcc tgcagatcac aggacttcga 600
gctgaggacg gtggcatctt ccactgtgtg gcctcaaaca tcgccagtat ccggatcagc 660
cacggggcca ggctcactgt gtcaggctcg ggctctgggg cctacaagga gccagccatc 720
ctcgtggggc ctgagaacct caccctgaca gtgcaccaga ccgcggtgct tgagtgtgtc 780
gccacgggca acccgcgccc cattgtgtcc tggagccgcc tggatggtcg ccctatcggg 840
gtggagggca tccaggtgct gggcacagga aacctcatca tctcagacgt gacggtccag 900
cactctggcg tctacgtctg tgcagccaac agacctggca cccgggtgag gagaacggca 960
cagggccggc tggtggtgca agccccagct gagtttgtgc agcatcccca gtccatctcc 1020
aggccagctg ggaccacagc catgttcacc tgccaagccc agggtgagcc accgcctcat 1080
gtcacgtggc tgaaaaatgg acaggtgctg gggccaggag gccacgtcag gctcaagaat 1140
aacaacagca cactgaccat ttctggaatc ggtcctgagg atgaagccat ttatcagtgt 1200
gtggccgaga acagtgcggg ctcatcacag gccagtgcca ggctgaccgt actgtgggct 1260
gaggggctcc ccgggcctcc ccgcaatgtg cgggcagtct ctgtgtcttc cactgaggtg 1320
cgtgtgtcct ggagtgagcc gctggccaac accaaggaga tcatcggcta cgtcctgcac 1380
atcaggaagg ctgctgaccc accggagctg gagtatcagg aggcagtcag caagagcacc 1440
tttcagcacc tggtcagcga cctggagccc tccacagcct acagtttcta catcaaggcc 1500

tacacaccaa gggggccag ctcagcctct gtgcccaccc tagctagcac cctgggtgaa 1560
gcccctgccc cacccccact gtcagtgcga gtcctgggca gctcctcctt gcagctgctg 1620
tgggagcctt ggccccggct ggcccagcac gagggcggct tcaagctgtt ttaccgccca 1680
gcaagcaaga cctccttcac cggccccatc ctgctgcctg gaaccgtctc ctcctacaac 1740
ctcagccagc tcgaccccac tgcagtgtat gaggtgaagc tgctcgccta caaccagcat 1800
ggagatggca atgccacagt ccgctttgtg tctttgaggg gagcatctga gaggacagcc 1860
ttgagcccac catgtgactg ccggaaggag gaggccgcca accagacgtc caccacaggc 1920
atcgtcatcg gcatccacat cggggtcact tgcatcatct tctgtgtcct cttcctcctg 1980
ttcggccaaa ggggcagggt cctcctgtgt aaagatgtgg aaaaccagct gtcccctcca 2040
cagggtcccc ggagccagag ggaccctggc attctggccc taaatggggc gagacgggga 2100
cagcggggcc agctgggccg agacgagaaa cgtgtggata tgaaggagct ggagcagctg 2160
ttccccccgg ccagcgcagc agggcagccg gaccccagac ccacagatcc tgcagccccc 2220
gctccgtgtg aggagaccca gctctccttg ctgccacttc aggggtgcgg cctgatggag 2280
gggaagacga cggaggcgaa gaccacagag gccacggctc cctgcgccgg cctggcggct 2340
gccccaccac ccccagatgg aggccctggc ctcctcagtg aaggccaggc ttccaggcct 2400
gcagcggccc gggttaccca gccagctcac tcggaacagt agccagtgtc tggcaggctc 2460
cagagggtgg acggagcggg gcccattctc aggtcaaaag caagatttct actgtcatgt 2520
gggatttgga tggtcctggg ggctccccag tatttctatc ctgactgcct cttgggttgt 2580
caaaacccaa ggcagccttg acagggaccc cccggcccta acacccatca ggagttggag 2640
cagttcctgc aggagcctgt tccttccctg ggctgacgcc cccttgcctc tgcctggtac 2700
ccacatgact tggaactgaa ctaacatttt tctttaaaaa gcaaaactta aaaaaaaaaa 2760
a 2761

<210> 42
<211> 1440
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 1420867CB1

<400> 42
agcagctcaa ccccatcttt ggagagatcc tggagctaag catctctctc ccagctgaga 60
cggagctgac ggtcgccata tttgatcatg acctcgtggg ttctgacgac ctcatcgggg 120
agacccacat tgatctggaa aaccgattct atagccacca cagagcaaac tgtgggctgg 180
cctcccagta tgaaagatgg ttacaatgcc tggcgtgatg cattctggcc ttcgcagatc 240
ctggcggggc tgtgccaacg ctgtggcctc cctgcccctg aataccgagc cggtgctgtc 300
aaggtgggca gcaaagtctt cctgacacca ccggagaccc tgcccccagg gatctcttca 360
catgtggatt gacatctttc ctcaagatgt gcctgctcca cccccagttg acatcaagcc 420
tcggcagcca atcagctatg agctcagagt tgtcatctgg aacacggagg atgtggttct 480
ggatgacgag aatccactca ccggagagat gtcgagtgac atctatgtga agagctgggt 540
gaaggggttg gagcatgaca agcaggagac agacgttcac ttcaactccc tgactgggga 600
ggggaacttc aattggcgct ttgtgttccg ctttgactac ctgcccacgg agcgggaggt 660
gagcgtccgg cgcaggtctg gacctttgc cctggaggag gcggagttcc ggcagcctgc 720
agtgctggtc ctgcaggatc cctggagttg cagctaccag acatggtgca tggggcccgg 780
ggccccgagc tctgctctgt gcagctggcc cgcaatgggg ccgggccgag gtgcaatctg 840
tttcgctgcc gccgcctgag gggctggtgg ccggtagtga agctgaagga ggcagaggac 900
gtggagcggg aggcgcagga ggctcaggct ggcaagaaga agcgaaagca gaggaggagg 960
aagggccggc cagaagacct ggagttcaca gacatgggtg gcaatgtgta catcctcacg 1020
ggcaaggtgg aggcagagtt tgagctgctg actgtggagg aggccgagaa acggccagtg 1080
gggaaggggc ggaagcagcc agagcctctg gagaaaccca gccgccccaa aacttccttc 1140
aactggtttg tgaacccgct gaagaccttt gtcttcttca tctggcgccg gtactggcgc 1200
accctggtgc tgctgctact ggtgctgctc accgtcttcc tcctcctggt cttctacacc 1260
atccctggcc agatcagcca ggtcatcttc cgtcccctcc acaagtgact ctcgctgacc 1320
ttggacactc acccagggtg ccaacccttc aatgcctgct cctggaagtc tttcttaccc 1380

atgtgagcta ccccagagtc tagtgcttcc tctgaataaa cctatcacag ccactgaaaa 1440

<210> 43
<211> 1066
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7512289CB1

<400> 43
agcagtctcc gccctcgcac gggagctggg aggctgcgag atccctaccg cagtagccgc 60
ctctgccgcc gcggagcttc ccgaacctct tcagccgccc ggagccgctc ccggagcccg 120
gccgtagagg ctgcaatcgc agccgggagc ccgcagcccg cgccccgagc ccgccgccgc 180
ccttcgaggg cgccccaggc cgcgccatgg tgaaggtgac gttcaactcc gctctggccc 240
agaaggaggc caagaaggac gagcccaaga gcggcgagga ggcgctcatc atccccccg 300
acgccgtcgc ggtggactgc aaggacccag atgatgtggt accagttggc caaagaagag 360
cctggtgttg gtgcatgtgc tttggactag catttatgct tgcaggtgtt attctaggag 420
gagcatactt gtacaaatat tttgcacttc aaccagatga cgtgtactac tgtggaataa 480
agtacatcaa agatgatgtc atcttaaatg agccctctgc agatgcccca gctgctctct 540
accagacaat tgaagaaaat attaaaatct ttgaagaaga agaagttgaa tttatcagtg 600
tgcctgtccc agagtttgca gatagtgatc ctgccaacat tgttcatgac tttaacaagg 660
ctggaaccta tttgcctcag tcctatctga ttcatgagca catggttatt actgatcgca 720
ttgaaaacat tgatcacctg ggtttcttta tttatcgact gtgtcatgac aaggaaactt 780
acaaactgca acgcagagaa actattaaag gtattcagaa acgtgaagcc agcaattgtt 840
tcgcaattcg gcattttgaa aacaaatttg ccgtggaaac tttaatttgt tcttgaacag 900
tcaagaaaaa cattattgag gaaaattaat atcacagcat aaccccaccc tttacatttt 960
gtgcagtgat tatttttaa agtcttcttt catgtaagta gcaaacaggg ctttactatc 1020
ttttcatctc attaattcaa ttaaaaccat taccttaaaa ttttaa 1066

<210> 44
<211> 4179
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7512447CB1

<400> 44
ggaggcgctc ggggcatcga ggctgggagg cgggtccgcc ccctattgtg tagcggcgag 60
agtggagccg agcggtgcgg agcagatctg gtggttctcc ggagagcagc ttcctcgggt 120
gttacatgag ccaagccctc actgtacaga agagtgagag ctgaaacctg ttccctgagc 180
tgatcagaag gacatcccct ggcccctcca tctgggctcc tgtggatagg aggggctggg 240
tgagcaggcc agctgggcta tggtgtggtg cctcggcctg gccgtcctca gcctggtcat 300
cagccagggg gctgacggtc gagggaagcc tgaggtggta tcggtggtgg gccgggctgg 360
ggagagtgtg gtgctgggct gtgacctgct gcccccggcc ggccggcccc ccctgcatgt 420
catcgagtgg ctgcgctttg gattcctgct tcccatcttc atccagttcg gcctctactc 480
tccccgaatt gaccctgatt acgtgggacg agtccggctg cagaaggggg cctctctcca 540
gattgagggt ctccgggtgg aagaccaggg ctggtacgag tgccgcgtgt tcttcctgga 600
ccagcacatc cctgaagacg attttgctaa cggctcctgg gtgcgtctga cagtcaattc 660
acccccctcaa ttccaggaga cacctcctgc tgtgttggaa gtgcaggaac tggagcctgt 720
gaccctgcgt tgtgtggccc gtggcagccc cctgcctcat gtgacgtgga agctccgagg 780
aaaggacctt ggccagggcc agggccaggt gcaagtgcag aacgggacgc tgcggatccg 840
ccgggtagag cgaggcagct ctggggtcta cacctgccaa gcctccagca ctgagggcag 900

cgccacccac gccacccagc tgctagtgct aggaccccca gtcatcgtgg tgccccccaa 960
gaacagcaca gtcaatgcct cccaggatgt ttcattggcc tgccatgctg aggcataccc 1020
tgctaacctc acctacagct ggttccagga caacatcaat gtcttccaca ttagccgcct 1080
gcagccccgg gtgcggatcc tggtggacgg gagcctgcgg ctgctggcca cccagcctga 1140
tgatgccggc tgctacacct gtgtgcccag caatggcctc ctgcatccac cctcagcctc 1200
tgcctacctc actgtgctct gcatgccggg ggtgatccgc tgcccggttc gtgccaaccc 1260
cccactgctc tttgtcagct ggaccaagga tggaaaggcc ctgcagctgg acaagttccc 1320
tggctggtcc cagggcacag aaggctcact gatcatcgcc ctggggaacg aggatgccct 1380
gggagaatac tcctgcaccc cctacaacag tcttggtacc gccgggccct ctcctgtgac 1440
ccgcgtgctg ctcaaggctc ccccagcttt tatagagcgg cccaaggaag aatatttcca 1500
agaagtaggg cgggagctgc tcatcccctg ctccgcccaa ggggaccctc ctcctgttgt 1560
ctcttggacc aaggtgggcc gggggctgca aggccaggcc caggtggaca gcaacagcag 1620
cctcatcctg cgaccattga ccaaggaggc ccacgggcac tgggaatgca gtgccagcaa 1680
tgctgtggcc cgagtggcca cctccacgaa cgtctacgtg ctgggcacta gccctcatgt 1740
tgtcaccaat gtgtccgtgg tggctttgcc caagggtgcc aatgtctcct gggagcctgg 1800
ctttgatggt ggttatctgc agagattcag tgtctggtac accccactgg ccaagcgtcc 1860
tgaccgaatg caccatgact gggtgtcctt ggcagtgcct gtgggggctg ctcacctcct 1920
agtgccaggg ctgcagcccc acacccagta ccagttcagc gtgctagctc agaacaagct 1980
ggggagtggt cccttcagcg aaatcgtctt gtctgctccg gaagggcttc ctaccacgcc 2040
agctgcaccc gggcttcccc caacagagat accgcctccc ctgtcccctc cgcggggtct 2100
ggtggcagtg aggacacccc ggggggtact cctgcattgg gatcccccag agctggtccc 2160
taagagactg gatggctacg tcttggaagg ccggcaaggc tcccagggct gggaggtgct 2220
ggacccggct gtggcaggca cagaaacaga gctgctggtg ccaggcctca tcaaggatgt 2280
tctctacgag ttccgcctcg tggccttcgc gggcagcttc gtcagcgacc ccagcaacac 2340
ggccaacgtc tccacttccg gtctggaggt ctacccttcg cgcacgcagc tgccgggcct 2400
cctgcctcag cccgtgctgg ccggcgtggt gggcggagtc tgctttctgg gagtggccgt 2460
ccttgtgagc atcctggccg gctgcctcct gaaccggcgc agggctgccc gccgccgccg 2520
caagcgcctc cgccaagatc cacctcttat cttctctccg accgggaagt cagctgcacc 2580
ctctgctctg ggctcaggca gtcctgacag cgtggcgaag ctgaagctcc agggatcccc 2640
agtccccagc ctgcgccaga gtctgctctg gggggatcct gccggaactc ccagccccca 2700
cccggatcct ccatctagcc ggggaccctt acctctggag cccatttgcc ggggcccaga 2760
cgggcgcttt gtgatggggc ccactgtggc ggccccccag gaaaggtcag gccgggagca 2820
ggcagaacct cggactccag cccagcgtct ggcccggtcc tttgactgta gcagcagcag 2880
ccccagtggg gcaccccagc ccctctgcat tgaagacatc agccctgtgg cacccctcc 2940
agcagcccca cccagtccct tgccaggtcc tggacccctg ctccagtacc tgagcctgcc 3000
cttcttccga gagatgaatg tggatgggga ctggcccccg cttgaggagc ccagccctgc 3060
tgcaccccca gattacatgg atacccggcg ctgtcccacc tcatctttcc ttcgttctcc 3120
agaaacccct cctgtatccc ccagggaatc acttcctggg gctgtggtag gggctggggc 3180
cactgcagag cccccttaca cagccctggc tgactggaca ctgagggagc ggctgctgcc 3240
aggccttctc cctgctgccc ctcgaggcag cctcaccagc cagagcagtg ggcgaggcag 3300
cgcttcgttc ctgcggcccc cctccacagc cccctctgca ggaggcagct acctcagccc 3360
tgctccagga gacaccagca gctgggccag tggccctgag agatggcccc gaagggagca 3420
tgtggtgaca gtcagcaaga ggaggaacac atctgtggac gagaactatg agtgggactc 3480
agaattccct ggggacatgg aattgctgga gactttgcac ctgggcttgg ccagctcccg 3540
gctcagacct gaagctgagc cagagctagg tgtgaagact ccagaggagg gctgcctcct 3600
gaacactgcc catgttactg gccctgaggc ccgctgtgct gcccttcggg aggaattcct 3660
ggccttccgc cgccgccgag atgctactag ggctcggcta ccagcctatc gacagccagt 3720
cccccacccc gaacaggcca ctctgctgtg aacatccctg atgtgaggct gtgaaaaggc 3780
atatggacct gcaaaggagg cccccaacca gacagaccta gtttcaaacg agggcactgc 3840
ccctgcctgc ccctttggtg cccaggcaca gaccctgata gtgggtttgg gtcaccttgg 3900
tatggaatgt atgtgctgac cccctaggtg agtctgggga ttggaacagg gatcttaggt 3960
ctgcctctct ctctctctct ctctctctct ctctctctgt gtgtgtgtgt gtgtgtgtga 4020
agtttttttac aggtgaataa cacaaagttt gaaagacaaa aaaaaaaagg gggggccgc 4080
gaatagtgag ctcgtcaccc cggaaataaa tcccggaccg gtcactgagg gtaaccagtc 4140
ccctataggg agtccgtata agagtcgcgg aacacaggg 4179

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US02/40059

A. CLASSIFICATION OF SUBJECT MATTER

IPC(7) : C07H 21/04; C12N 15/00, 15/85, 15/86; C07K 5/00, 14/00; A61K 38/00

US CL : 435/69.1,325; 514/2; 530/300,350; 536/23.1

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 435/69.1,325; 514/2; 530/300,350; 536/23.1

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

Please See Continuation Sheet

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	US 2002/0155577 A1 (KOUTNIKOVA et al.) 24 October 2002 (24.10.2002), entire document.	1-7, 9-10, 12-13, 56, 69, 78, 91
A	WANG, X. Rim1 and Raphpilin-3 Bind Rab3-GTP by Composite Determinants Partially Related through N-terminal alpha-Helix Motifs. The Journal of Biological Chemistry 31 August 2001, Vol. 276, No. 35, pages 32480-32488, entire document.	1-7, 9-10, 12-13, 56, 69, 78, 91
A	STROM, M. A Family of Rab27-binding Proteins. The Journal of Biological Chemistry 12 July 2002, Vol. 277, No. 28, pages 25423-25430, entire document.	1-7, 9-10, 12-13, 56, 69, 78, 91
A	BURGOYNE, R.D. Control of Membrane Fusion Dynamics during Regulated Exocytosis. Biochemical Society Tranactions August 2001, Vol. 29, No. 4, pages 467-472, entire document	1-7, 9-10, 12-13, 56, 69, 78, 91
A	FUKUDA et al. Novel Splicing Isoforms of Synaptotagmin-like Proteins 2 and 3: identification of the Slp Homology domain. Biochemical and Biophysical Research Communications 04 May 2001, Vol. 283, No. 2, pages 513-519, entire document.	1-7, 9-10, 12-13, 56, 69, 78, 91
X	WO 01/46256 A2 (INCYTE GENOMICS, INC.) 28 June 2001 (28.06.2001), pages 28-29 (SEQ ID NO: 26 shares 62.3% homology with SEQ ID NO: 23).	1-7, 9-10, 12-13, 56, 69, 78, 91

☒ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search

25 April 2003 (25.04.2003)

Date of mailing of the international search report

23 MAY 2003

Name and mailing address of the ISA/US

Commissioner of Patents and Trademarks
Box PCT
Washington, D.C. 20231

Facsimile No. (703)305-3230

Authorized officer

Christopher Nichols, Ph.D.

Telephone No. 703-308-0196

Form PCT/ISA/210 (second sheet) (July 1998)

INTERNATIONAL SEARCH REPORT

C. (Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 01/60857 A2 (AVENTIS PHARMA S.A.) 23 August 2001 (23.08.2001), Claim 10, Figure 9 (87.7% homology with SEQ ID NO: 1), Claim 10 pages 77-80 (87.7% homology with SEQ ID NO: 1); and pages 1-27 (SEQ ID NO: 1 shares 58.8% homology with SEQ ID NO: 23; SEQ ID NO: 12 shares 68.1% homology with SEQ ID NO: 23; SEQ ID NO: 14 shares 62.5% homology with SEQ ID NO: 23; SEQ ID NO: 42 shares 68.1% homology with SEQ ID NO: 23; SEQ ID NO: 44 shares 62.5% homology with SEQ ID NO: 23).	1-7, 9-10, 12-13, 56, 69, 78, 91
X	WO 01/75067 A2 (HYSEQ, INC.) 11 October 2001 (11.10.2001), claim 20, SEQ ID NO: 47280 (shares 71.1% homology with SEQ ID NO:1), SEQ ID NO: 47279 (shares 77.9% homology with SEQ ID NO: 1).	1-7, 9-10, 12-13, 56, 69, 78, 91
X	WANG et al. Novel Rapphilin-3-like Protein Associates with Insulin-Containing Granules in Pancreatic Beta Cells. The Journal of Biological Chemistry 01 October 1999, Vol. 274, No. 40, pages 28543-28548 (Figure 1, 36.5% homology with SEQ ID NO: 1).	1-7, 9-10, 12-13, 56, 69, 78, 91
X	CASTETS et al. Zinedin, SG2NA, and Striatin Are Calmodulin-binding, WD Repeat Proteins Principally Expressed in the Brain. The Journal of Biological Chemistry 30 June 2000, Vol. 275, No. 26, pages 19970-19977 (Figure 1; 94.7% and 90.3% homology with SEQ ID NO: 14).	1-7, 9-10, 12-13, 56, 69, 78, 91
Y	HAYNES et al. A Direct Inhibitory Role for the Rab3-specific Effector, Noc2, in Ca2+-regulated Exocytosis in Neuroendocrine Cells. The Journal of Biological Chemistry, Vol. 276, No. 13, pages 9726-9732 (Figure 3).	1-7, 9-10, 12-13, 56, 69, 78, 91
Y	MATESIC et al. Mutations in Mlph, encoding a member of the Rab effector family, cause the melanosome transport defects observed in leaden mice. PNAS 28 August 2001, Vol. 98, No. 18, pages 10238-10243 (Figure 4).	1-7, 9-10, 12-13, 56, 69, 78, 91

INTERNATIONAL SEARCH REPORT

BOX II. OBSERVATIONS WHERE UNITY OF INVENTION IS LACKING

This application contains the following inventions or groups of inventions which are not so linked as to form a single general inventive concept under PCT Rule 13.1. In order for all inventions to be examined, the appropriate additional examination fees must be paid.

Group 1, claim(s) 1-7, 9-10, 12-13, 17-18, 56, and 78 (each in part), drawn to a method of producing a polypeptide comprising SEQ ID NO: 1, an isolated polypeptide comprising SEQ ID NO: 1, an isolated polynucleotides comprising SEQ ID NO: 23, cells transformed with said polynucleotides, promoter sequences, and compositions comprising same.

Group 2, claim(s) 1-7, 9-10, 12-13, 17-18, 57, and 79 (each in part), drawn to a method of producing a polypeptide comprising SEQ ID NO: 2, an isolated polypeptide comprising SEQ ID NO: 2, an isolated polynucleotides comprising SEQ ID NO: 24, cells transformed with said polynucleotides, promoter sequences, and compositions comprising same.

Group 3 claim(s) 1-7, 9-10, 12-13, 17-18, 58, and 80 (each in part), drawn to a method of producing a polypeptide comprising SEQ ID NO: 3, an isolated polypeptide comprising SEQ ID NO: 3, an isolated polynucleotides comprising SEQ ID NO: 25, cells transformed with said polynucleotides, promoter sequences, and compositions comprising same.

Group 4 claim(s) 1-7, 9-10, 12-13, 17-18, 59, and 81 (each in part), drawn to a method of producing a polypeptide comprising SEQ ID NO: 4, an isolated polypeptide comprising SEQ ID NO: 4, an isolated polynucleotides comprising SEQ ID NO: 26, cells transformed with said polynucleotides, promoter sequences, and compositions comprising same.

Group 5 claim(s) 1-7, 9-10, 12-13, 17-18, 60, and 82 (each in part), drawn to a method of producing a polypeptide comprising SEQ ID NO: 5, an isolated polypeptide comprising SEQ ID NO: 5, an isolated polynucleotides comprising SEQ ID NO: 27, cells transformed with said polynucleotides, promoter sequences, and compositions comprising same.

Group 6 claim(s) 1-7, 9-10, 12-13, 17-18, 61, and 83 (each in part), drawn to a method of producing a polypeptide comprising SEQ ID NO: 6, an isolated polypeptide comprising SEQ ID NO: 6, an isolated polynucleotides comprising SEQ ID NO: 28, cells transformed with said polynucleotides, promoter sequences, and compositions comprising same.

Group 7 claim(s) 1-7, 9-10, 12-13, 17-18, 62, and 84 (each in part), drawn to a method of producing a polypeptide comprising SEQ ID NO: 7, an isolated polypeptide comprising SEQ ID NO: 7, an isolated polynucleotides comprising SEQ ID NO: 29, cells transformed with said polynucleotides, promoter sequences, and compositions comprising same.

Group 8 claim(s) 1-7, 9-10, 12-13, 17-18, 63, and 85 (each in part), drawn to a method of producing a polypeptide comprising SEQ ID NO: 8, an isolated polypeptide comprising SEQ ID NO: 8, an isolated polynucleotides comprising SEQ ID NO: 30, cells transformed with said polynucleotides, promoter sequences, and compositions comprising same.

Group 9 claim(s) 1-7, 9-10, 12-13, 17-18, 64, and 86 (each in part), drawn to a method of producing a polypeptide comprising SEQ ID NO: 9, an isolated polypeptide comprising SEQ ID NO: 9, an isolated polynucleotides comprising SEQ ID NO: 31, cells transformed with said polynucleotides, promoter sequences, and compositions comprising same.

Group 10 claim(s) 1-7, 9-10, 12-13, 17-18, 65, and 87 (each in part), drawn to a method of producing a polypeptide comprising SEQ ID NO: 10, an isolated polypeptide comprising SEQ ID NO: 10, an isolated polynucleotides comprising SEQ ID NO: 32, cells transformed with said polynucleotides, promoter sequences, and compositions comprising same.

Group 11, claim(s) 1-7, 9-10, 12-13, 17-18, 66, and 88 (each in part), drawn to a method of producing a polypeptide comprising SEQ ID NO: 11, an isolated polypeptide comprising SEQ ID NO: 11, an isolated polynucleotides comprising SEQ ID NO: 33, cells transformed with said polynucleotides, promoter sequences, and compositions comprising same.

Group 12, claim(s) 1-7, 9-10, 12-13, 17-18, 67, and 89 (each in part), drawn to a method of producing a polypeptide comprising SEQ ID NO: 12, an isolated polypeptide comprising SEQ ID NO: 12, an isolated polynucleotides comprising SEQ ID NO: 34, cells transformed with said polynucleotides, promoter sequences, and compositions comprising same.

Group 13 claim(s) 1-7, 9-10, 12-13, 17-18, 68, and 90 (each in part), drawn to a method of producing a polypeptide comprising SEQ ID NO: 13, an isolated polypeptide comprising SEQ ID NO: 13, an isolated polynucleotides comprising SEQ ID NO: 35, cells transformed with said polynucleotides, promoter sequences, and compositions comprising same.

Group 14 claim(s) 1-7, 9-10, 12-13, 17-18, 69, and 91 (each in part), drawn to a method of producing a polypeptide comprising SEQ ID NO: 14, an isolated polypeptide comprising SEQ ID NO: 14, an isolated polynucleotides comprising SEQ ID NO: 36, cells transformed with said polynucleotides, promoter sequences, and compositions comprising same.

Group 15 claim(s) 1-7, 9-10, 12-13, 17-18, 70, and 92 (each in part), drawn to a method of producing a polypeptide comprising SEQ ID NO: 15, an isolated polypeptide comprising SEQ ID NO: 15, an isolated polynucleotides comprising SEQ ID NO: 37, cells transformed with said polynucleotides, promoter sequences, and compositions comprising same.

Group 16 claim(s) 1-7, 9-10, 12-13, 17-18, 71, and 93 (each in part), drawn to a method of producing a polypeptide comprising SEQ ID NO: 16, an isolated polypeptide comprising SEQ ID NO: 16, an isolated polynucleotides comprising SEQ ID NO: 38, cells transformed with said polynucleotides, promoter sequences, and compositions comprising same.

Group 17 claim(s) 1-7, 9-10, 12-13, 17-18, 72, and 94 (each in part), drawn to a method of producing a polypeptide comprising SEQ ID NO: 17, an isolated polypeptide comprising SEQ ID NO: 17, an isolated polynucleotides comprising SEQ ID NO: 39, cells transformed with said polynucleotides, promoter sequences, and compositions comprising same.

Group 18 claim(s) 1-7, 9-10, 12-13, 17-18, 73 and 95 (each in part), drawn to a method of producing a polypeptide comprising SEQ ID NO: 18, an isolated polypeptide comprising SEQ ID NO: 18, an isolated polynucleotides comprising SEQ ID NO: 40, cells transformed with said polynucleotides, promoter sequences, and compositions comprising same.

Group 19 claim(s) 1-7, 9-10, 12-13, 17-18, 74, and 96 (each in part), drawn to a method of producing a polypeptide comprising SEQ ID NO: 19, an isolated polypeptide comprising SEQ ID NO: 19, an isolated polynucleotides comprising SEQ ID NO: 41, cells transformed with said polynucleotides, promoter sequences, and compositions comprising same.

Group 20 claim(s) 1-7, 9-10, 12-13, 17-18, 75 and 97 (each in part), drawn to a method of producing a polypeptide comprising SEQ ID NO: 20, an isolated polypeptide comprising SEQ ID NO: 20, an isolated polynucleotides comprising SEQ ID NO: 42, cells transformed with said polynucleotides, promoter sequences, and compositions comprising same.

Group 21, claim(s) 1-7, 9-10, 12-13, 17-18, 76, and 98 (each in part), drawn to a method of producing a polypeptide comprising SEQ ID NO: 21, an isolated polypeptide comprising SEQ ID NO: 21, an isolated polynucleotides comprising SEQ ID NO: 43, cells transformed with said polynucleotides, promoter sequences, and compositions comprising same.

Group 22, claim(s) 1-7, 9-10, 12-13, 17-18, 77, and 99 (each in part), drawn to a method of producing a polypeptide comprising SEQ ID NO: 22, an isolated polypeptide comprising SEQ ID NO: 22, an isolated polynucleotides comprising SEQ ID NO: 44, cells transformed with said polynucleotides, promoter sequences, and compositions comprising same.

Group 23 claim(s) 8, drawn to a transgenic organism.

Group 24 claim(s) 11, 31, 32, 34, and 36-43 drawn to a method of preparing an antibody, polyclonal and monoclonal antibodies, and compositions comprising same.

Group 25 claim(s) 14-16, drawn to a method of detecting a target polynucleotides in a sample.

Group 26 claim(s) 19 and 22, drawn to a method for treating a disease or condition associated with *decreased* expression of functional NTRN.

Group 27 claim(s) 20 and 21, drawn to a method of screening for effectiveness as an *agonist* of a polypeptide and compositions comprising same.

Group 28 claim(s) 23 and 24, drawn to drawn to a method of screening for effectiveness as an *antagonist* of a polypeptide and compositions comprising same.

Group 29 claim(s) 25, drawn to drawn to a method for treating a disease or condition associated with *overexpression* of functional NTRN.

Group 30 claim(s) 26, drawn to a method of screening for a compound that specifically binds to a polypeptide.

INTERNATIONAL SEARCH REPORT

Group 31, claim(s) 27, drawn to a method of screening for a compound that modulates the activity of a polypeptide.

Group 32, claim(s) 28, drawn to a method of screening a compound for effectiveness in altering expression of a target polynucleotide.

Group 33 claim(s) 29, drawn to a method of assessing toxicity of a test compound.

Group 34 claim(s) 30 and 35, drawn to a method for a diagnostic test for a condition or disease associated with the expression of NTRAN in a biological sample.

Group 35 claim(s) 44, drawn to a method of detecting a polypeptide.

Group 36 claim(s) 45, drawn to a method of purifying a polypeptide.

Group 37 claim(s) 46 and 48-55, drawn to a microarray.

Group 38 claim(s) 47, drawn to a method of generating an expression profile of a sample.

The inventions listed as Groups 1-38 do not relate to a single general inventive concept under PCT Rule 13.1 because, under PCT Rule 13.2, they lack the same or corresponding special technical features for the following reasons:

Group 1 recites the special technical feature of SEQ ID NO: 1, which is not required by any of the other groups.

Continuation of B. FIELDS SEARCHED Item 3:
WEST (PG Pubs, USPT, JPO,EPO,DERWENT); NCBI (PubMed); STN (CAPLUS,MEDLINE)
human neurotransmission-associated protein, SNAP, SNARE, synaptotagmin, Rab, WD Repeat, vescile, neurotrasmitter, SEQ ID NO: 1, 14, 23, 36

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US02/40059

Box I Observations where certain claims were found unsearchable (Continuation of Item 1 of first sheet)

This international report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claim Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. ☐ Claim Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

3. ☐ Claim Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of Item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:
Please See Continuation Sheet

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.
2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.
3. ☒ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.: 1-7, 9-10, 12-13, 17-18, 56, 69, 78, and 91 (each in part)

4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest
☐ The additional search fees were accompanied by the applicant's protest.
☐ No protest accompanied the payment of additional search fees.